

Hovenga, Nina; Bos, Wilfried  
**Bildungsmonitoring auf der Systemebene**

2011, 57 S.



Quellenangabe/ Reference:

Hovenga, Nina; Bos, Wilfried: Bildungsmonitoring auf der Systemebene. 2011, 57 S. - URN: urn:nbn:de:0111-pedocs-107509 - DOI: 10.25656/01:10750

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-107509>

<https://doi.org/10.25656/01:10750>

### Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

### Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

### Kontakt / Contact:

**peDOCS**  
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation  
Informationszentrum (IZ) Bildung  
E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

Mitglied der

  
Leibniz-Gemeinschaft

---

# Bildungsmonitoring auf der Systemebene

Nina Hovenga  
Wilfried Bos



---

## UDiKom

**Aus- und Fortbildung der Lehrkräfte  
in Hinblick auf Verbesserung der  
Diagnosefähigkeit, Umgang mit  
Heterogenität, individuelle Förderung**

Die Produktion dieses Materials  
zum Einsatz in die Lehrerbildung  
wurde ermöglicht durch

Deutsche Telekom Stiftung





## **Bildungsmonitoring auf der Systemebene**

Alle im Projekt erstellten Materialien  
finden Sie unter  
[www.udikom.de](http://www.udikom.de)



**Inhaltsverzeichnis – Teil 3 – Bildungsmonitoring auf der Systemebene**

<b>3.0 Einleitung</b>	3
<b>3.1 Zielsetzung</b>	5
3.1.1 Begründung für Systemmonitoring-Studien	5
3.1.2 Ziele von Systemmonitoring-Studien	6
3.1.3 Vorstellung von verschiedenen Systemmonitoring-Studien	7
3.1.3.1 TIMSS (optional)	8
3.1.3.2 PISA (optional)	8
3.1.3.3 IGLU (optional)	10
3.1.3.4 DESI (optional)	11
3.1.4 Weiterführende Literatur	12
3.1.5 Verständnis-Aufgaben und Diskussionspunkte	12
<b>3.2 Bewertungskriterien im Systemmonitoring</b>	13
3.2.1 Beschreibung der Bezugsnormen	13
3.2.1.1 Kriteriale Bezugsnorm	13
3.2.1.2 Soziale Bezugsnorm	13
3.2.1.3 Individuelle Bezugsnorm	14
3.2.2 ‚Literacy‘-Orientierung vs. curriculare Orientierung	14
3.2.3 Kompetenz, Kompetenzniveaus, Kompetenzmodelle und Standards	17
3.2.3.1 Kompetenz	17
3.2.3.2 Kompetenzmodelle	18
3.2.3.3 Kompetenzniveaus	18
3.2.3.4 Standards	20
3.2.4 Weiterführende Literatur	20
3.2.5 Verständnis-Aufgaben und Diskussionspunkte	20
<b>3.3 Testkonstruktion</b>	21
3.3.1 Testtheoretischer Hintergrund	21
3.3.2 Testgütekriterien und Testanforderungen	22
3.3.3 Test-Verfahrensweisen	22
3.3.4 Testaufgaben Auswahl	23
3.3.5 Bezugsgruppen	26
3.3.6 Übersicht durchgeführter und anstehender Erhebungen (optional)	27
3.3.7 Weiterführende Literatur	28
3.3.8 Verständnis-Aufgaben und Diskussionspunkte	28
<b>3.4 Anwendungsbereich</b>	29
3.4.1 Theoretische Grundlagen	29
3.4.2 Wesentliche Ergebnisse (optional)	29
3.4.2.1 TIMSS (optional)	30
3.4.2.2 PISA (optional)	33
3.4.2.3 IGLU (optional)	39
3.4.2.4 DESI (optional)	41
3.4.3 Verwendung der Ergebnisse	42
3.4.4 Weiterführende Literatur	43
3.4.5 Verständnis-Aufgaben und Diskussionspunkte	43
<b>3.5 Praktische Implikationen</b>	44
3.5.1 Rezeption der Studien (optional)	45
3.5.2 Entscheidungen auf Grund der Ergebnisse	46
3.5.2.1 Sieben Handlungsfelder der KMK nach PISA 2000	47
3.5.2.2 Neue Schwerpunkte der KMK	47
3.5.3 Auswirkungen des Systemmonitorings auf die Schulpraxis	48
3.5.3.1 Standardentwicklung	48
3.5.3.2 Einführung von Kernlehrplänen	49
3.5.3.3 Stärkung der Sprachkompetenz	50
3.5.3.4 Ganztagschulen	51
3.5.3.5 SINUS / SINUS transfer	51
3.5.3.6 For.mat	52
3.5.3.7 UDiKom	52
3.5.3.8 Bildungsberichterstattung	52
3.5.4 Weiterführende Literatur	53
3.5.5 Verständnis-Aufgaben und Diskussionspunkte	53
<b>3.6 Literaturverzeichnis</b>	54

### 3.0 Einleitung

Der vorliegende Studienbrief-Teil beschäftigt sich mit dem Systemmonitoring als Teil des Bildungsmonitorings<sup>1</sup>. Dabei wird unter Bildungsmonitoring in der Literatur verstanden:

#### Definition: Bildungsmonitoring

„Bildungsmonitoring meint dabei die kontinuierliche und systematische, auf wissenschaftliche Methoden gestützte Beobachtung der Bedingungen, Verläufe, Ergebnisse und Wirkungen von Bildungsprozessen in und außerhalb von Institutionen mit dem Ziel, bildungspolitischen Akteuren wissenschaftlich aufbereitetes, empirisches Wissen zur Information für bildungspolitische Entscheidungsprozesse zur Verfügung zu stellen.“ (Wolter, 2009)

„Bildungsmonitoring ist die systematische und auf Dauer angelegte Beschaffung und Aufbereitung von Informationen über ein Bildungssystem und dessen Umfeld.“ (Maritzen, 2008)

Im Jahr 2006 wurde durch die KMK eine Gesamtstrategie zum Bildungsmonitoring beschlossen. Bildungsprozesse in Deutschland sollen durch vier miteinander verbundene Bereiche beobachtet und weiterentwickelt werden:

- regelmäßige Teilnahme an internationalen Schulleistungsuntersuchungen
- zentrale Überprüfung der Bildungsstandards im Ländervergleich
- Vergleichsarbeiten
- gemeinsame Bildungsberichterstattung

Das übergeordnete Ziel dieser Gesamtstrategie liegt in der Beschaffung von Informationen, die für die Steuerung des Bildungssystems und für die Schul- und Unterrichtsentwicklung benötigt werden. Zudem sollen die gewonnenen Erkenntnisse mit Maßnahmen der Qualitätsentwicklung verknüpft werden, sodass die pädagogische Arbeit an jeder einzelnen Schule profitieren kann (KMK, 2006).

Wichtig für die Durchführung des Bildungsmonitorings ist es, dass Wissenschaft, Bildungspolitik und Bildungspraxis miteinander arbeiten (Stadelmann, 2008). Durch die kombinierte Arbeit von Theorie und Praxis können Studien im Rahmen des Bildungsmonitorings nicht nur ‚bildungspolitischen Akteuren‘ sondern auch Lehrkräften bei ihrer Arbeit behilflich sein. Zudem gibt es viele Berührungspunkte einer Lehrkraft mit dem Bildungsmonitoring. Aus diesen Punkten folgt, dass es auch für Lehrkräfte von Bedeutung ist, über die Studien des Systemmonitorings Bescheid zu wissen, da sie einen Einfluss auf den Lehralltag haben (werden).

Hierzu zählen:

- Diskussionen in den Medien über die Wirksamkeit von Schule
- Diskussionen innerhalb des Kollegiums
- Eltern, die durch die erhöhte Transparenz der Schulwirksamkeit wissen wollen, ob ihre eigenen Kinder besser oder ähnlich abgeschnitten hätten
- Konsequenzen der Bildungsadministration:
  - Bildungsstandards (vgl. Kapitel 3.5.3.1)
  - Kernlehrpläne (vgl. Kapitel 3.5.3.2)
  - Sprachförderung (vgl. Kapitel 3.5.3.3)
  - Ganztagschulen (vgl. Kapitel 3.5.3.4)
- Projekte (KMK, Ministerien, Schulamt, Stadt) (vgl. Kapitel 3.5.3.5 bis 3.5.3.7)
- eigene Schülerinnen und Schüler, Schülerinnen und Schüler der eigenen Schule oder Schülerinnen und Schüler von bekannten Lehrkräften werden getestet

Von großer Bedeutung ist es, den Begriff des ‚Bildungsmonitorings‘ von dem der ‚Evaluation‘ abzugrenzen. Zwar werden auch im Rahmen des Bildungsmonitorings evaluative Aussagen über Teilpopulationen, beispielsweise im Vergleich der Bundesländer oder der Schulformen getroffen, dennoch soll der Begriff ‚Evaluation‘ dafür reserviert werden, dass der „Erfolg einer bestimmten Maßnahme bzw. die Leistungsfähigkeit einer einzelnen Institution (z.B. einer Schule) zu beurteilen ist“ (Klieme u.a., 2007). Die weiteren Unterschiede fasst Tabelle 1 zusammen.

Bildungs-  
monitoring /  
Evaluation

<sup>1</sup> Im Folgenden wird die Bezeichnung ‚Systemmonitoring als Teil des Bildungsmonitorings‘ verkürzt als ‚Systemmonitoring‘ verwendet.

<sup>2</sup> Die Bedeutungen dieser Abkürzungen werden in den Kapiteln 3.1.3.1 bis 3.1.3.4 erklärt.

	Bildungsmonitoring	Evaluation
Grundlage/Ziel	Informationen über das Bildungssystem	Erfolg bestimmter Maßnahmen/Institutionen
Untersuchungsgegenstand	Unterschiedliche Bereiche des Bildungssystems	Bestimmte Maßnahme
Untersuchungsart	Umfassend	Eng und detailliert
Ergebnis	Informationen	Konkrete Entscheidung

Tabelle 1: Abgrenzung der Begriffe ‚Bildungsmonitoring‘ und ‚Evaluation‘

Durch die Unterschiede wird deutlich, dass eine Studie nicht gleichzeitig beiden ‚Anliegen‘ gerecht werden kann (Klieme u.a., 2007). Allerdings ist es möglich, im Rahmen des ‚normalen‘ Bildungsmonitorings ergänzende Befragungen mit evaluativem Charakter durchzuführen. Beispielsweise wurden in einigen Studien ergänzende Videoanalysen erstellt, um das Lehrverhalten zu betrachten, andere Studien betrachteten den Zusammenhang zwischen den Ergebnissen und dem sozialen Hintergrund der Schülerinnen und Schüler. Zudem wurden in den meisten Studien Zusatzstudien auf nationaler Ebene durchgeführt, sodass hierdurch untersucht werden konnte, ob beispielsweise Bildungsstandards erreicht werden (Klieme u.a., 2007).

Der vorliegende Studienbrief-Teil befasst sich mit einem Teil des Bildungsmonitorings, dem Systemmonitoring.

### Definition: Systemmonitoring

Systemmonitoring ist ein Teil des Bildungsmonitorings und bezeichnet „Maßnahmen der quantitativen Erfassung von Bildungserträgen auf Systemebene“ (Stanat, 2008).

### 3.1 Zielsetzung

Kapitel 3.1 beschäftigt sich mit der Zielsetzung von Systemmonitoring. Dazu wird betrachtet, warum Studien des Systemmonitorings durchgeführt werden und welche Ziele sie im Einzelnen verfolgen. Zudem werden die vier Studien TIMSS<sup>2</sup>, PISA, IGLU und DESI vorgestellt, die bekanntesten Studien im Rahmen des Systemmonitorings.

In diesem Kapitel werden folgende Fragen beantwortet:

- Welche Charakteristika haben Systemmonitoring-Studien?
- Warum werden Systemmonitoring-Studien durchgeführt?
- Welche Studien gibt es und welche Inhalte haben diese?

Bei den vier oben genannten Studien handelt es sich jeweils um Large-Scale Assessments.

#### Definition: Large-Scale Assessments

- Leistungsmessungen mit einer hohen Anzahl an Testpersonen
- Untersuchungen, die messen, welche Kompetenzen Personen einer bestimmten Zielgruppe haben
- internationaler Vergleich dieser Kompetenzen
- Stichprobenuntersuchungen, bei denen nur in einigen wenigen Schulen aus jedem Staat Schülerinnen und Schüler getestet werden
- Erhebungen in vielen verschiedenen Staaten
- Durchführung sehr genau geplant, in der Praxis nicht sehr flexibel

Eine Rückmeldung der Ergebnisse an die an den Studien beteiligten Schulen ist nur eingeschränkt möglich. Vor allem die Vorgehensweise der Rückmeldung ist als nicht adressatengerecht kritisiert worden (Rolff, 2008), zudem erfolgt sie durch die komplexe und internationale Auswertung der Daten erst sehr spät. Auch ist die Bedeutung der Datenrückmeldung an die einzelnen Schulen, die an den Erhebungen teilgenommen haben, durch die Vergleichsarbeiten (vgl. Studienbrief-Teil Vergleichsarbeiten) zurückgegangen, da diese alle Klassen und alle Schülerinnen und Schüler betreffen.

Daten-  
rückmeldung  
an Schulen

Eine solche Rückmeldung an die Schulen ist aber auch nicht das bedeutsamste Ziel von Systemmonitoring-Studien.

#### 3.1.1 Begründung für Systemmonitoring-Studien

Systemmonitoring ist sinnvoll und notwendig, da nur so Informationen für die Steuerung des Bildungssystems gewonnen werden können. Nur auf diese Weise können sich Bildungsinstitutionen entwickeln und optimieren.

Wichtigkeit  
Systemmoni-  
toring

Abbildung 1 zeigt diesen Verlauf.

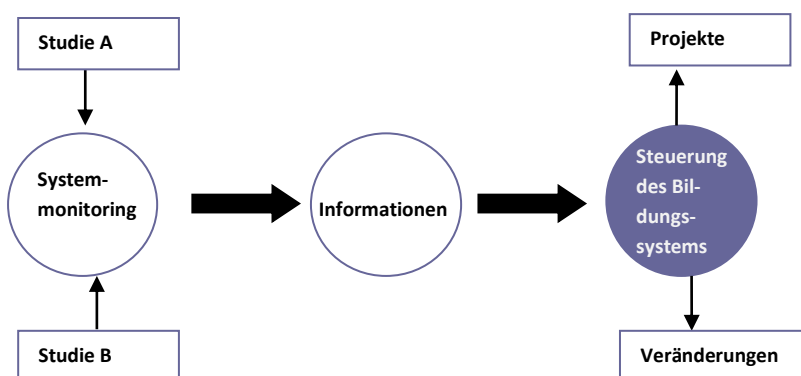


Abbildung 1: Funktionsweise des Systemmonitorings

Die Abbildung zeigt, dass sich das Systemmonitoring aus mehreren Projekten (hier: A und B) zusammensetzt. Aus diesen Studien ergeben sich dann Informationen, die für die Steuerung des Bildungssystems verwendet werden, beispielsweise um Projekte zu initiieren und Veränderungen zu bewirken.

<sup>2</sup> Die Bedeutungen dieser Abkürzungen werden in den Kapiteln 3.1.3.1 bis 3.1.3.4 erklärt.



Um zu verdeutlichen, warum das Systemmonitoring wichtig für den Schulalltag ist, zeigt Abbildung 2 den Einfluss auf den Unterricht, der zwischen der Lehrkraft und der individuellen Förderung der Schülerinnen und Schüler steht. Das Ziel des Unterrichts ist es, dass der Schüler individuell gefördert wird und einen Wissenszuwachs erreicht. Dazu erhält er einen Unterricht, der durch die Lehrkraft gestaltet wird. Aber nicht nur die Lehrkraft hat einen Einfluss auf den Unterricht, sondern auch die Projekte und Veränderungen, die sich aus der Steuerung des Bildungssystems ergeben.

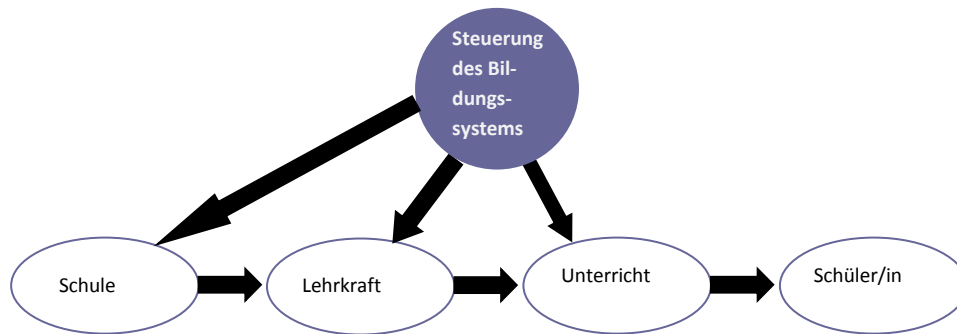


Abbildung 2: Einfluss des Systemmonitorings auf den Alltag einer Lehrkraft

Anders als beispielsweise in den USA gab es in Deutschland bis zum Anfang der 1990er Jahre keine Informationen darüber, wie erfolgreich schulische Bildungsprozesse in Deutschland verlaufen. Bildungsmonitoring fand bis zu diesem Zeitpunkt nicht oder nur unzureichend statt. Wie Bildungsmonitoring funktioniert, war bis zu diesem Zeitpunkt weder der Öffentlichkeit noch großen Teilen der Wissenschaft bekannt (Köller, 2008). Stattdessen fand in Deutschland bis vor einigen Jahren eine „reine Input-Steuerung“ statt, der Output dagegen wurde nicht betrachtet.

Standardisierten Leistungsmessungen stand man in Deutschland lange Zeit kritisch gegenüber, sie wurden als „Übergriff der Bildungsverwaltung, als Angriff auf die professionelle Verantwortung der Lehrkräfte empfunden“ (Klieme u.a., 2007). Zudem ging man davon aus, dass vergleichbare Klassen auch vergleichbare Lernfortschritte machen und deutsche Schülerinnen und Schüler im internationalen Vergleich gut abschneiden (KMK, 2006). Zwar wurde in der Wissenschaft viel über Qualitätsentwicklung diskutiert, die Qualitätssicherung dagegen wurde nicht angegangen (Klieme/Baumert, 2001).

Den Ausschlag dafür, dass das Bildungssystem intensiv und regelmäßig untersucht wird, gaben die Ergebnisse der Studie TIMSS im Jahr 1995. Durch die Rezeption der Ergebnisse wurden nicht nur inhaltliche Probleme des Bildungssystems deutlich, sondern auch, in welchem Rahmen sich Veränderungen ergeben könnten (Specht, 2008). Um Probleme des Schulsystems schneller erkennen und beheben zu können, finden Erhebungen im Rahmen des Bildungsmonitoring seitdem regelmäßig statt (vgl. Übersicht in Kapitel 3.3.6).

Durch die erste Teilnahme an internationalen Schulleistungsuntersuchungen wurde den deutschen bildungspolitischen Akteuren immer deutlicher, wie Bildungsmonitoring wirken kann und wo die Vorteile liegen. Daher entschied sich die Ständige Konferenz der Kultusminister der Länder (KMK) im Oktober 1997 in ihren ‚Konstanzer Beschlüssen‘ dazu, regelmäßiges Bildungsmonitoring in Deutschland zu implementieren. Von diesem Zeitpunkt an wurde die Qualitätssicherung im Bildungswesen zum zentralen Thema erklärt. Dazu gehörte die Teilnahme an internationalen Schulleistungstudien und die Einführung von Bildungsmonitoring auf nationaler Ebene (PISA-Konsortium Deutschland, 2007). Zu den Aktivitäten auf nationaler Ebene zählt beispielsweise eine regelmäßige Bildungsberichterstattung (vgl. Kapitel 3.5.3.8).

Aufbauend auf diesem Beschluss und den folgenden Ergebnissen der Systemmonitoring-Studien entwickelte die KMK im Jahr 2002 sieben Handlungsfelder und setzte im Jahr 2006 nach weiteren Studien neue Schwerpunkte. Diese Handlungsfelder und Schwerpunkte sollen als erste Handlungen auf Grund des Bildungsmonitorings verstanden und daher in Kapitel 3.3.5 ‚Praktische Implikationen‘ näher betrachtet werden.

### 3.1.2 Ziele von Systemmonitoring-Studien

Systemmonitoring-Studien verfolgen gemeinsam mit einer kontinuierlichen Bildungsberichterstattung und einer verstärkten Förderung der Bildungsforschung das Ziel zu erfahren, ob und wo das Bildungssystem optimiert werden muss (Stanat, 2008). Das Ziel der Studien ist es, „auf einer breiten empirischen Basis die Beschreibung und Analyse der Erträge fachlichen Lernens in den Mittelpunkt“ zu rücken (Köller, 2008). Auf Grund der Durchführung und der Interpretation der Studien ergibt sich eine erhöhte Transparenz bildungspolitischer Entscheidungsprozesse, dadurch kann Handlungsbedarf deutlich gemacht werden (Stanat, 2008). Zudem kann überprüft werden, ob Reformmaßnahmen, die auf Grund von Ergebnissen vorangegangener Studien umgesetzt wurden, ihre Wirkung entfalten können (Klieme u.a., 2007).

Die Studien dienen auch dazu, empirisch abgesicherte Befunde darüber zu erhalten, in welchen (Kompetenz)-Bereichen die Schülerinnen und Schüler Stärken und Schwächen haben. Eine solche Information ist wichtig, damit „die Bildungschancen aller gewahrt werden“ können und eine erhöhte Betrachtung der individuellen Voraussetzungen der Schülerinnen und Schüler stattfinden kann (KMK, 2005). Einige Studien untersuchen auch die Einstellungen und Motive der Schülerinnen und Schüler. Dadurch ergeben sich Informationen, die, gemeinsam mit den Informationen über die Kompetenzen, bei der Entscheidung über Schulstrukturen, Lehrpläne, die Lehrerbildung und die Schulbuchgestaltung von großer Relevanz sein können (Klieme u.a., 2001).

Abgesehen von diesen übergeordneten Zielen verfolgen die einzelnen Studien aber auch eigene Ziele. Als Beispiel sollen hier die Ziele der DESI-Studie zitiert werden:

„Die DESI-Studie soll erstmals aussagefähige Daten zu den Fähigkeiten deutscher Schülerinnen und Schüler im aktiven und passiven Gebrauch des Deutschen und Englischen erheben, dokumentieren, Erklärungsmodelle unter Einschluss personaler, unterrichtlicher und schulischer Faktoren ableiten und Optimierungsansätze für den Unterricht aufzeigen.“ (Eichler, 2003)

Ein weiteres Ziel von Systemmonitoring-Studien, das vor allem in der Öffentlichkeit dankbar aufgenommen wird, ist die Bereitstellung von so genannten ‚benchmarks‘ (Klieme u.a., 2001).

### Definition: benchmarks

Der Begriff ‚benchmarks‘ beschreibt in diesem Kontext Leistungsergebnisse sehr erfolgreicher Bildungssysteme im In- und Ausland. Mit diesen Leistungsergebnissen kann dann das ‚eigene‘ Ergebnis verglichen werden. Dabei muss im Rahmen des Systemmonitorings darauf geachtet werden, dass die Ergebnisse erst durch Zusammenhanganalysen erklärt und anschließend Stärken und Schwächen eines Systems benannt werden können. Zu beachten ist, dass es im Systemmonitoring nicht ausschließlich um ein reines ‚benchmarking‘, geschweige denn um ein ‚ranking‘ der teilnehmenden Bildungssysteme geht (Klieme u.a., 2001) (siehe weitere Ziele in Kapitel 1.2.).

Auch in Zukunft wird sich Deutschland dem internationalen Vergleich stellen, damit die „erforderliche Anschlussfähigkeit“ gesichert werden kann. Zudem können so die entwickelten Bildungsstandards (s. Kapitel 3.5.2.1) an internationale Maßstäbe angelehnt und überprüft werden (KMK, 2006).

### 3.1.3 Vorstellung von verschiedenen Systemmonitoring-Studien

Viele Aspekte haben alle Studien des Systemmonitorings gemeinsam. Diese sollen hier zunächst erläutert werden. Im Anschluss daran besteht die Möglichkeit, sich über die einzelnen Studien näher zu informieren.

- Sie testen international eine bestimmte Kompetenz von Schülerinnen und Schülern einer festgelegten Altersgruppe.
- Die Untersuchung findet zeitgleich und mit denselben Aufgaben in allen teilnehmenden Staaten statt.
- Die Federführung liegt bei einer allgemein anerkannten Institution, die die Durchführung und Auswertung der Daten plant und begleitet.
- In einigen Teilnehmerstaaten werden zusätzliche Kompetenzen innerhalb der Studien getestet um einen intranationalen Vergleich zu ermöglichen.
- Die Federführung dieser Zusatz- oder Ergänzungsstudien liegt im jeweiligen Land selbst.

In Hinblick auf die Ergebnisse der Studien ist zu beachten: Studien wie PISA geben keine Antwort darauf, an welcher Stelle und auf welche Weise die gemessenen Kompetenzen erworben wurden und auch nicht, wie und wo die gemessenen Unterschiede in den Kompetenzen entstanden sind. Dafür fehlen notwendige Kontrollgruppen und die Studien sind nicht als Längsschnittstudien angelegt. Allerdings kann durch Systemmonitoring-Studien deutlich werden, ob und wo es Zusammenhänge zwischen sozialer Herkunft und Bildungserfolg gibt. Dies wird durch den „systemvergleichenden Charakter der internationalen Leistungsstudien“ möglich (Specht, 2008).

Keine Antworten von  
System-  
monitoring-  
Projekten

### Definition: Längsschnittstudien

Längsschnittstudien beschreiben Studien, die eine Situation oder Kompetenz nicht nur zu einem Zeitpunkt untersuchen. In Längsschnittstudien werden die Testpersonen zu mindestens zwei Zeitpunkten getestet, sodass sich eine zeitliche Entwicklung darstellen lässt. In Deutschland werden Längsschnittstudien im Rahmen des Nationalen Bildungspanels (NEPS) (siehe <http://www.uni-bamberg.de/neps/>) durchgeführt.

Im Folgenden sollen die vier Studien TIMSS, PISA, IGLU und DESI kurz zusammengefasst vorgestellt werden. Es handelt sich nicht um eine komplette Darstellung der einzelnen Studien. Für einen Überblick über bereits durchgeführte und noch anstehende Studien bietet Kapitel 3.3.6 zwei Tabellen, die diese darstellen.

**3.1.3.1 TIMSS (optional)<sup>3</sup>**

TIMSS stand zunächst für ‚Third International Mathematics and Science Study‘ und ist die Nachfolgeuntersuchung zu FIMS und SIMS (First and Second International Mathematics Study) und FISS und SISS (First and Second International Science Study), die allerdings ohne deutsche Beteiligung durchgeführt wurden. Die Zielsetzung von TIMSS liegt darin, die Mathematik- und Naturwissenschaftsleistungen in der Grundschule (TIMSS I), in der Sekundarstufe I (TIMSS II) und in der Sekundarstufe II (TIMSS III<sup>4</sup>) gleichzeitig zu untersuchen. Deutschland beteiligte sich 1995 nur an TIMSS II und III, 2007 nur an TIMSS I.

Die Studie wurde ab 1999 mit dem Namen ‚Trends in International Mathematics and Science Study‘ alle vier Jahre durchgeführt, wobei sich Deutschland im weiteren Verlauf nur noch 2007 beteiligte.

Ergänzungsstudie

Im Jahr 1995 wurde zusätzlich n die Studie ‚TIMSS-Deutschland‘ durchgeführt. In dieser Ergänzungsstudie wurden Schülerinnen und Schüler aus den Klassen 7 und 8 zusätzlich in einer Längsschnittstudie untersucht. Hier wurde ein Schwerpunkt auf die motivationale Entwicklung und die verwendeten Unterrichtsmethoden gelegt. 2007 fand eine solche deutsche Ergänzungsstudie nicht statt.

Tabelle 2 zeigt eine Übersicht über die teilnehmenden Schülerzahlen in Deutschland:

Teilnehmende

	TIMSS I	TIMSS II	TIMSS III
1995	----	7.000 Schülerinnen und Schüler	5.000 Schülerinnen und Schüler
2007	5.496 Schülerinnen und Schüler	----	----

Tabelle 2: Deutsche Teilnehmerzahlen TIMSS 1995 und 2007

International nahmen an TIMSS 2007 425.000 Schülerinnen und Schüler aus 59 Staaten teil.

TIMSS ist aber nicht eine reine Untersuchung der Kompetenzen der Schülerinnen und Schüler, sondern beschäftigt sich, beispielsweise im Jahr 1995, mit fünf Bereichen:

- International vergleichende Analysen von Lehrplänen und Lehrbüchern
- Schulleistungsuntersuchungen in den mathematisch-naturwissenschaftlichen Fächern
- Befragungen der Schulleiter und Fachlehrer
- Videoaufnahmen im Mathematikunterricht in Deutschland, Japan und den USA
- Ethnographische Fallstudien in Deutschland, Japan und den USA

Inhaltlich wurden die drei Bereiche Mathematik (Arithmetik, Geometrie/Messen, Daten), Naturwissenschaften (Biologie, Physik, Geographie) und kognitive Anforderungsbereiche (Reproduzieren, Anwenden, Problemlösen) untersucht (Baumert / Lehmann, 1997; Baumert / Bos / Watermann, 1998; Bos / Bonsen / Baumert / Prenzel / Selter / Walther, 2008).

**3.1.3.2 PISA (optional)**

PISA ist mit Sicherheit die bekannteste Studie des Systemmonitorings. Die Abkürzung steht für ‚Programme for International Student Assessment‘. PISA ist Teil des Indikatorenprogramms der OECD (Organisation für wirtschaftliche Zusammenarbeit und Entwicklung), „dessen Ziel es ist, den OECD-Mitgliedsstaaten vergleichende Daten über die Ressourcenausstattung, die individuelle Nutzung sowie die Funktions- und Leistungsfähigkeit ihrer Bildungssysteme zur Verfügung zu stellen“. Zudem werden auch Zusammenhänge zwischen den betrachteten Kompetenzen und der sozialen und kulturellen Herkunft der Schülerinnen und Schüler weltweit betrachtet.

Untersuchungsgegenstand

Untersucht werden in PISA Grundbildungsaspekte in den drei Bereichen Lesekompetenz, mathematische Grundbildung und naturwissenschaftliche Grundbildung. Dabei wird immer in jedem Jahr einer der drei Aspekte als Schwerpunkt betrachtet. Tabelle 3 gibt eine Übersicht über die Schwerpunkte.

<sup>3</sup> Optionale Kapitel dienen der Vertiefung und können bei Bedarf genutzt werden.

<sup>4</sup> Beachtet werden muss dabei, dass in Deutschland in TIMSS III auch das berufliche Schul- und Ausbildungswesen mit betrachtet wurde, sodass auch Auszubildende in Teilzeit-berufsschulen Teil der getesteten Schülerinnen und Schüler waren. Auch in den anderen teilnehmenden Staaten wurden in TIMSS III jeweils unterschiedliche Bildungsgänge untersucht.

	Lesekompetenz	Mathematische Grundbildung	Naturwissenschaftliche Grundbildung
2000	X		
2003		X	
2006			X
2009	X		
2011		X	
2014			X

Tabelle 3: Schwerpunktgebiete in PISA

„Grundbildungsaspekte“ bedeutet dabei, dass in PISA Kompetenzen erfasst werden, die „in modernen Gesellschaften für eine befriedigende Lebensführung in persönlicher und wirtschaftlicher Hinsicht sowie für eine aktive Teilnahme am gesellschaftlichen Leben notwendig“ sind.

### Beispiel: Aufgabe aus PISA 2000 - PIZZA

Eine Pizzeria bietet zwei runde Pizzas mit der selben Dicke in verschiedenen Größen an. Die kleinere hat einen Durchmesser von 30 cm und kostet 30 Zeds. Die größere hat einen Durchmesser von 40 cm und kostet 40 Zeds.

Beispielaufgabe 1: Mathematische Grundbildung

Bei welcher Pizza bekommt man mehr für sein Geld? Gib eine Begründung an.<sup>5</sup>

Es wird deutlich, dass durch Systemmonitoring-Studien nicht sinnfreies Wissen abgefragt wird, sondern das Wissen immer mit der Anwendung in der Praxis verknüpft wird.

Als Beispiel dafür sollen die konkreten Fragestellungen von PISA 2000 dargestellt werden:

### Beispiel: PISA 2000: „Wie gut können Schülerinnen und Schüler ...

... geschriebene Texte unterschiedlicher Art in ihren Aussagen, Absichten und ihrer formalen Struktur verstehen, einordnen und sachgerecht nutzen?“

... verständnisvoll mit Mathematik umgehen und mathematische Werkzeuge in einer Vielfalt von Kontexten einsetzen?“

... naturwissenschaftliches Wissen anwenden und Schlussfolgerungen ziehen, welche die natürliche Welt und durch menschliches Handeln in ihr vorgenommene Veränderung betreffen?“

... fächerübergreifende Basiskompetenzen (selbstreguliertes Lernen und Vertrautheit mit Computern) einsetzen?“

Dabei wird vor allem beim letzten Punkt deutlich, dass neben den fachspezifischen Kompetenzen auch fachübergreifende Kompetenzen (so genannte ‚cross-curricular competencies‘) wie der Umgang mit dem Computer untersucht wurden.

Auf internationaler Ebene wurden zudem biografische Hintergründe und Familien- und Lebensverhältnisse der Schülerinnen und Schüler erfasst, um unterstützende Informationen für die Suche nach Begründungen der Ergebnisse zu erhalten.

Tabelle 4 gibt eine Übersicht über die teilnehmenden Staaten und die Anzahl der teilnehmenden Schülerinnen und Schüler:

Teilnehmende

	International: Schülerinnen und Schüler	Teilnehmende Staaten	Deutsche Schülerinnen und Schüler
2000	180.000	32	5.000
2003	250.000	41	4.660
2006	400.000	56	4.891
2009	470.000	65	4.979

Tabelle 4: Übersicht über an PISA teilnehmende Schülerinnen, Schüler und Staaten

<sup>5</sup> Weitere PISA-Beispielaufgaben werden auf [http://pisa.ipn.uni-kiel.de/fr\\_reload.html?beispielaufgaben.html](http://pisa.ipn.uni-kiel.de/fr_reload.html?beispielaufgaben.html) gesammelt vorgestellt.

Ergänzungsstudie

Die internationale Studie wurde um eine nationale Untersuchungen unter dem Namen PISA-E („E“ für Ergänzung) erweitert. Eine solche Erweiterung ist in allen Systemmonitoring-Studien problemlos möglich, so lange es keine Reibungspunkte mit der internationalen Untersuchung gibt. Diese könnten beispielsweise darin liegen, dass die Ergänzungsstudien zu viel Zeit und Aufwand einnehmen und die teilnehmenden Schülerinnen und Schüler nicht mehr genügend Zeit und Konzentration für die internationale Studie haben.

Die Ergänzungsstudien gehen mit einer Vergrößerung der Stichprobe einher. Dieses Vorgehen ist notwendig, damit „aussagekräftige regionale Vergleiche innerhalb der Länder auf der Basis einer ausreichend großen Stichprobe“ ermöglicht werden können (van Ackeren, 2006).

Durch eine nationale Untersuchung können einerseits sowohl die einzelnen Bundesländer als auch die Schulformen untereinander verglichen werden.

Ziel der Ergänzungsstudie

Ziel von PISA-E ist es, Stärken und Schwächen der Schülerinnen und Schüler in den Bundesländern herauszuarbeiten. Durch diesen Vergleich können die Leistungen der einzelnen Bundesländer in Beziehung zum deutschen Durchschnitt, zum OECD-Durchschnitt und auch zueinander gesetzt und analysiert werden. Dabei soll vermieden werden, dass PISA-E zu einer „Bildungsolympiade“ wird, es geht nicht um Gewinner und Verlierer, sondern um den positiven Vergleich und die Übernahme erfolgreicher Konzepte (vgl. den Begriff „benchmarks“ in Kapitel 3.1.1).

Zudem wurden noch weitere Untersuchungen im Rahmen von PISA durchgeführt. Als Beispiel soll hier PISA 2003 dienen. Hier wurde unter dem Namen PISA-I-Plus (wobei das „I“ für die internationale Untersuchung steht) eine Längsschnittstudie durchgeführt. Dafür wurden weitere Schülerinnen und Schüler ausgewählt und über den Zeitraum eines gesamten Schuljahres untersucht.

PISA aktuell

Die letzte Erhebung von PISA fand im Sommer 2009 statt. An PISA 2009 nahmen international rund 470 000 Schülerinnen und Schüler aus 65 Staaten (mit allen OECD-Staaten) teil. Im Vergleich zur ersten Erhebung im Jahr 2000 mit 43 Teilnehmerstaaten, sowie zu den Erhebungen in den Jahren 2003 mit 41 Teilnehmerstaaten, und 2006 mit 57 Teilnehmerstaaten lässt sich eine deutliche Steigerung erkennen. (Deutsches PISA-Konsortium 2001; PISA-Konsortium Deutschland 2005, 2006, 2007, 2008, 2010; Baumert u.a. 2002; Kiper/Kattmann 2003; Kiper 2003a)

### 3.1.3.3 IGLU (optional)

Nach den Ergebnissen von TIMSS 1995 wurde deutlich, dass eine Beteiligung Deutschlands im Grundschulbereich hilfreich gewesen wäre, da so hätte festgestellt werden können, welche der in TIMSS aufgetretenen Defizite der Schülerinnen und Schüler bereits aus der Grundschulzeit resultieren. Daher kam die Möglichkeit gelegen, sich an IGLU zu beteiligen. IGLU steht für „Internationale Grundschul-Lese-Untersuchung“ und untersucht international die Lesekompetenz von Grundschülerinnen und Grundschulern. Sie fand zum ersten Mal im Jahr 2001 statt. Im internationalen Rahmen wird die Studie „Progress in International Reading Literacy Study“ (PIRLS) genannt. Geplant ist eine regelmäßige Durchführung alle fünf Jahre, sodass auch im Jahr 2006 Daten im Rahmen von IGLU erhoben wurden. Hier wurden „mittels authentischer Texte<sup>6</sup> verschiedener Textgattungen unterschiedliche Aspekte der Kompetenz im Rahmen verschiedener Leseabsichten“ untersucht, zudem wurden „die Fähigkeit zum Schreiben und die kognitiven Lernvoraussetzungen ermittelt“ (Döbert et al., 2009). Die nächste Untersuchung findet im Jahr 2011 statt.

In Deutschland wurden 2001 zusätzlich zum Leseverständnis an einem zweiten Testtag auch die Bereiche Mathematik, Naturwissenschaften, Orthographie und Aufsatz im Rahmen der Studie IGLU-E („E“ für Ergänzung) untersucht. In den Bereichen Mathematik und Naturwissenschaften konnten Test-Items aus TIMSS (s. Kapitel 3.1.3.1) verwendet werden, sodass die Ergebnisse der getesteten Schülerinnen und Schüler nachträglich auf der internationalen TIMSS-Skala eingeordnet werden konnten. Für den Bereich Orthographie wurde ein Lückentext verwendet, in dem Wörter diktiert wurden, die dem Grundwortschatz von Viertklässlern entsprechen sollten. 2001 beteiligten sich allerdings nur 12 der 16 Bundesländer an dieser zusätzlichen Untersuchung, im Jahr 2006 nahmen alle 16 Bundesländer teil.

Auch in dieser Studie wurden Fragebögen an Eltern, Lehrer, Schulleiter und Schülerinnen und Schüler verteilt und somit viele Zusatzinformationen erhoben, um Informationen für eine Interpretation der Ergebnisse und einen Einblick in die Lebenswelt der Schülerinnen, Schüler, Lehrer, Eltern und Schulleiter zu erhalten.

Tabelle 5 gibt eine Übersicht über die an IGLU/PIRLS teilnehmenden Staaten und die Anzahl der Schülerinnen und Schüler:

Einordnung auf TIMSS-Skala

<sup>6</sup> Authentische Texte bezeichnen Texte, die in der Realität verwendet wurden oder theoretisch verwendet werden könnten.

	Internationale Schülerinnen und Schüler	Teilnehmende Staaten bzw. Regionen	Deutsche Schülerinnen und Schüler
<b>2001</b>	146.490	35	10.571 <sup>7</sup>
<b>2006</b>	215.137	45	8.302 <sup>8</sup>

Tabelle 5: Übersicht teilnehmende Staaten, Schülerinnen und Schüler an IGLU/PIRLS

Für die Studie wurden in allen Ländern Schülerinnen und Schüler aus der Klassenstufe getestet, die die meisten neunjährigen Kinder umfasste. In den meisten Teilnehmerstaaten, so auch in Deutschland, entspricht dies der Jahrgangsstufe 4. Ausgenommen waren aus der Zufallsauswahl solche Klassen, in denen die Kinder aufgrund geistiger oder körperlicher Behinderung nicht zur Bearbeitung der Aufgaben in der Lage waren (vgl. Kapitel 3.3.5).

(Bos u.a., 2003, 2007, 2008)

### 3.1.3.4 DESI (optional)

An DESI nahmen Schülerinnen und Schüler der Jahrgangsstufe 9 teil. Die Abkürzung DESI steht für ‚Deutsch Englisch Schülerleistungen International‘. Die Studie wurde im Jahr 2001 von der KMK in Auftrag gegeben, um die „sprachlichen Leistungen und die Unterrichtswirklichkeit in den Fächern Deutsch und Englisch“ zu untersuchen. Das ‚I‘ in DESI steht für die Teilnahme der Länder Schweiz und Österreich, zudem sollten Aspekte des Erwerbs der Fremdsprache Englisch international verglichen werden. Ziel von DESI war, die Leistungen der Schülerinnen und Schüler sowohl in der ersten Fremdsprache Englisch als auch in der Unterrichtssprache Deutsch in den Bereichen ‚aktiver Sprachgebrauch‘, ‚Sprachbewusstheit‘ und ‚Kommunikation‘ zu untersuchen (vgl. Beschreibung des Ziels von DESI in Kapitel 3.1.2). Dabei wurden schriftliche und mündliche Kompetenzen und somit alle Teilkompetenzen und alle Lernbereiche des Unterrichts erfasst.

Untersuchungs-  
gegenstand

Der Begriff der ‚Unterrichtssprache‘ wird hier gewählt, da die Studie davon ausgeht, dass ein großer Anteil der Schülerinnen und Schüler Deutsch nicht als Erstsprache spricht. Daher wurden alle Schülerinnen und Schüler in der Auswertung danach kategorisiert, ob sie ‚Deutsch als Erstsprache‘ oder ‚Deutsch als Zweitsprache‘ sprechen oder ob sie bilingual aufgewachsen sind.

Unterrichts-  
sprache

DESI kann somit als Ergänzung zu PISA verstanden werden, da dort nur die Lesekompetenz der Schülerinnen und Schüler in der Unterrichtssprache getestet wurde.

Am Anfang und am Ende des Schuljahres 2003/2004 wurden in Deutschland 10.639 Schülerinnen und Schüler der neunten Klassen aller Schulformen getestet (insgesamt 219 Schulen, davon 40 mit bilinguaalem Zweig). Somit handelt es sich bei DESI um eine Längsschnittstudie (vgl. Definition in Kapitel 3.1.3). Zusätzlich zu den Schülerinnen und Schülern wurden die Lehrer, Eltern und Schulleitungen befragt und es wurden Videoaufnahmen im Englischunterricht durchgeführt. Durch die Anlage der Studie konnten bundesweit repräsentative Ergebnisse erzielt und durch die zusätzlichen Befragungen Erkenntnisse über Lehr- und Lernprozesse und den sprachlichen Kompetenzerwerb gewonnen werden. Diese Informationen sind sowohl für das allgemeine Systemmonitoring als auch für die Praxis von großer Bedeutung.

Durchführung

Die von DESI getesteten Teilkompetenzen werden in der folgenden Tabelle dargestellt:

getestete  
Kompetenzen

	Deutsch	Englisch
<b>Hörverstehen</b>		X
<b>Leseverständnis</b>	X (auch: Lesegeschwindigkeit)	X (global)
<b>Wortschatz</b>	X	
<b>Grammatik</b>	X	
<b>Sprachbewusstheit</b>	X (Grammatik, Stil und Rechtschreibung)	X (Grammatik und Adressatengerechtigkeit)
<b>Mündliche Sprachkompetenz</b>	X (Kommunikation/Argumentation)	X
<b>Rechtschreibung (Satzkonstruktion, Orthographie, Zeichensetzung)</b>	X	
<b>Schreiben</b>	X	X (kreativ)
<b>Interkulturelle Kompetenz</b>		X
<b>Globale Sprachkompetenz</b>		X

Tabelle 6: DESI-Teilkompetenzen in Deutsch und Englisch

<sup>7</sup> Nur 7.633 Schülerinnen und Schüler erfüllten die Anforderungen des Zufallkriteriums, sodass nur diese Anzahl in die Auswertung eingegangen ist.

<sup>8</sup> Einige Schülerinnen und Schüler waren am Testtag nicht anwesend, andere hatten keine Genehmigung der Eltern, sodass letztendlich 7.899 Schülerinnen und Schüler aus Deutschland teilgenommen haben.



Auch wenn sich diese Studie in einigen Bereichen deutlich von den anderen Studien des Systemmonitorings unterscheidet, ist ihre Betrachtung wichtig, da die Ergebnisse (vgl. Kapitel 3.4.2.4) gezeigt haben, dass „eine zentrale Ursache für die unbefriedigenden Ergebnisse im Unterrichtsgeschehen selbst zu suchen ist“ und als Konsequenz daraus mit der Einführung der Bildungsstandards (vgl. Kapitel 3.5.3.1) versucht wurde, dieser Schwäche entgegenzuwirken (Köller, 2008).

(Klieme, 2006; Klieme/Baumert, 2001; Eichler, 2003, Beck/Klieme, 2007; DESI-Konsortium, 2008)

### 3.1.4 Weiterführende Literatur

Eine Übersicht zum Thema ‚Bildungsmonitoring‘ bietet die Dokumentation zum OECD/CERI-Regionalseminar für die deutschsprachigen Länder in Potsdam („Bildungsmonitoring, Vergleichsstudien und Innovation“, Hg: LISUM, mn:ukk, EDK, 2008). Detaillierte Angaben über die Charakteristika der einzelnen Studien bieten jeweils die Veröffentlichungen der Studien. Zudem vermittelt das Buch ‚Leistungsmessungen in Schulen‘ von Franz E. Weinert (2001) sehr ausführliche Informationen über das Thema.

Beispielaufgaben können im Internet unter den folgenden Links gefunden werden:

PISA: [http://pisa.ipn.uni-kiel.de/fr\\_reload.html?beispielaufgaben.html](http://pisa.ipn.uni-kiel.de/fr_reload.html?beispielaufgaben.html)

TIMSS: <http://timss.ifs-dortmund.de/31.html>

### 3.1.5 Verständnis-Aufgaben und Diskussionspunkte

1. *Eine Schule hat an PISA teilgenommen und ein Ergebnis ähnlich dem finnischen erreicht. Der Schulleiter lädt zu einer Pressekonferenz mit der Überschrift: „Wir sind besser als Finnland“. Welche Fragen könnten die Journalisten ihm stellen?*
2. *Hinterfragen Sie die Entscheidung, bei PISA ein bestimmtes Alter und nicht eine Jahrgangsstufe untersucht zu haben.*
3. *In welchen Bereichen des Studiendesigns unterscheidet sich die PISA-Studie von der DESI-Studie und welche Bedeutung hat das für die Interpretation der Ergebnisse?*
4. *Formulieren Sie Gründe, warum ein Land nicht an einer (Teil-)Erhebung teilnehmen möchte/teilnimmt/teilnehmen kann und suchen Sie Argumente dafür, dass sich immer mehr Staaten im Laufe der Zeit dazu entschieden haben, beispielsweise an PISA teilzunehmen.*

## 3.2 Bewertungskriterien im Systemmonitoring

Kapitel 3.2 beschreibt die drei Bezugsnormen sozial, kriterial und individuell. Diese Bezugsnormen werden ebenfalls in den Studienbrief-Teilen ‚Individuelle Diagnostik‘ und ‚Vergleichsarbeiten‘ beschrieben, in Kapitel 3.2 allerdings mit Bezug auf das Systemmonitoring und die damit verbundenen Studien. Zudem wird auf die Unterscheidung zwischen einer curricularen und einer ‚literacy‘-Orientierung eingegangen. Es schließen sich Ausführungen über Kompetenzen, Kompetenzmodelle und Kompetenzniveaus und deren Zusammenhang mit der Standardentwicklung an.

In diesem Kapitel werden folgende Fragen beantwortet:

- *Wie werden die drei Bezugsnormen (kriterial, sozial, individuell) in den Systemmonitoring-Studien angewendet?*
- *Was bezeichnet der Begriff ‚literacy‘?*
- *Worin unterscheidet sich eine ‚literacy‘-Orientierung von einer curricularen Orientierung?*
- *Was sind Kompetenzmodelle und Kompetenzniveaus, wie werden sie erstellt und wie funktionieren sie?*
- *Was sind Standards und wie hängen sie mit Kompetenzmodellen und Kompetenzniveaus zusammen?*

### 3.2.1 Beschreibung der Bezugsnormen

Für Leistungsvergleiche werden Maßstäbe beziehungsweise ein Referenzrahmen benötigt. Drei mögliche Bezugsnormen kommen in Frage: ‚kriterial‘, ‚sozial‘ und ‚individuell‘. Da es sich bei den Studien des Systemmonitorings um Large-Scale Assessments (vgl. Definition in Kapitel 3.1) handelt, werden nicht Testergebnisse einzelner Schülerinnen und Schüler sondern Testergebnisse einzelner Staaten miteinander verglichen.

#### 3.2.1.1 Kriteriale Bezugsnorm

Kriteriale Bezugsnorm bedeutet, dass ein objektiv definiertes Kriterium als Vergleichsmaßstab dient und eine bestimmte Kompetenz erreicht werden muss (Näheres zum Begriff Kompetenz in Kapitel 3.2.3.1 und in Studienbrief-Teil Vergleichsarbeiten).

Die benötigten Kriterien können auf unterschiedliche Art gewonnen werden. Eine Möglichkeit ist der Rückgriff auf die Lehrpläne. Bei einer internationalen Studie ist dazu die Erarbeitung eines ‚weltweiten Curriculums‘ notwendig, so ist beispielsweise bei TIMSS vorgegangen worden. Die Alternative ist eine Orientierung an Grundkompetenzen, auf die sich die beteiligten Wissenschaftlerinnen und Wissenschaftler verständigen.

Die Orientierung an einer kriterialen Bezugsnormen ermöglicht außerdem eine differenzierte Auswertung, die über die reine Feststellung von Mittelwerten hinausgeht, in dem z.B. angegeben werden kann, wie viele Schülerinnen und Schüler welches der als Kriterium definierten Kompetenzniveaus erreicht haben.

Da aber eine bloße Aussage über das (Nicht-)Erreichen bestimmter Kompetenzen beziehungsweise Standards nicht alle Ziele der Systemmonitoring-Studien abdecken kann (vgl. die Ziele der Studien in Kapitel 3.1.2), muss auch die soziale Bezugsnorm betrachtet werden. Insbesondere TIMSS und PISA zeigen, wie eine solche Verbindung funktionieren kann (Klieme u.a., 2007).

#### 3.2.1.2 Soziale Bezugsnorm

Eine soziale Bezugsnorm besteht, wenn ein Ergebnis mit den Ergebnissen einer Referenzgruppe verglichen wird. In den Systemmonitoring-Studien existiert diese soziale Bezugsnorm, da die Teilnehmerstaaten und in den Ergänzungsstudien auch die einzelnen Bundesländer verglichen werden. Somit kann diese Bezugsnorm im Rahmen des Bildungsmonitorings auch ‚nationale‘ beziehungsweise ‚internationale‘ Bezugsnorm genannt werden.

IGLU zeigt einen sinnvollen Umgang mit Vergleichen nach sozialen Bezugsnormen: Generell bietet die Studie die Möglichkeit, sich mit jedem teilnehmenden Staat zu vergleichen. Allerdings ist es sinnvoller, sich mit Staaten zu vergleichen, „die einen ähnlichen wirtschaftlichen und kulturellen Hintergrund haben“. Nur so kann abgeschätzt werden, inwieweit die Leistungen der Schülerinnen und Schüler „als ein Ergebnis des Einsatzes von Ressourcen und kultureller Erfahrung zu betrachten und zu bewerten“ sind (Bos u.a., 2003). Die folgende Tabelle stellt die ‚IGLU-Ländervergleichsgruppen‘ dar.

Vergleich mit  
‚ähnlichen‘  
Ländern



Vergleichsgruppe 1 <sup>9</sup>	Vergleichsgruppe 2 <sup>10</sup>		Vergleichsgruppe 3 <sup>11</sup>
Deutschland	Bulgarien	Niederlande	Argentinien
England	Deutschland	Norwegen	Belize
Frankreich	England	Rumänien	Iran
Griechenland	Frankreich	Schottland	Kolumbien
Italien	Griechenland	Schweden	Kuweit
Niederlande	Island	Slowakei	Marokko
Schottland	Italien	Slowenien	Mazedonien
Schweden	Kanada	Tschechien	Türkei
	Lettland	Ungarn	
	Litauen	USA	
	Neuseeland	Zypern	

Tabelle 7: Ländervergleichsgruppen bei IGLU (Bos u.a., 2003)

Die drei Vergleichsgruppen dienen dazu, dass neben dem internationalen Mittelwert auch die Mittelwerte der drei Vergleichsgruppen als weitere Kennwerte für die Analyse angegeben werden können und so eine bessere Einschätzung über das eigene Abschneiden möglich ist. Es handelt sich also um eine Art ‚fairen Vergleich‘ (vgl. Studienbrief-Teil Vergleichsarbeiten).

In den Bildungsmonitoring-Studien ist sowohl die soziale als auch die kriteriale Bezugsnorm von großer Bedeutung.

### 3.2.1.3 Individuelle Bezugsnorm

Bei der ‚individuellen‘ oder ‚verlaufsorientierten‘ Bezugsnorm wird durch wiederholte Einsetzung eines Tests in der gleichen Probandengruppe die Veränderung der Kompetenzen betrachtet.

Bei den Untersuchungen des Systemmonitorings ist diese Betrachtungsweise auf Grund der jeweils neu gezogenen Stichprobe und der wechselnden Untersuchungsschwerpunkte nicht sinnvoll.

Allerdings kann eine individuelle Vergleichbarkeit dargestellt werden, wenn jedes Land als ein ‚Individuum‘ angesehen wird. Dadurch, dass die Studien in regelmäßigen Abständen durchgeführt werden, können jedoch beispielsweise die Ergebnisse von PISA 2000 und PISA 2003 innerhalb der einzelnen Staaten verglichen werden. Dabei ist aber zu beachten, dass bei PISA jeweils ein Schwerpunktthema verstärkt untersucht wird und somit ein intensiver Vergleich nur alle neun Jahre möglich ist. Dies funktioniert allerdings nur bei gleichbleibender Rahmenkonzeption. Über eine Veränderung dieser wird vor jeder Erhebung von den Mitgliedsstaaten entschieden. Aus einer Veränderung der Rahmenkonzeption folgt dann, dass bei der Betrachtung der Trends Einschränkungen entstehen können, da durch eine Konzeptionsveränderung teilweise andere Kompetenzen untersucht werden.

Durch einen Vergleich nach individueller Bezugsnorm werden Leistungssteigerungen oder Leistungsrückgänge der einzelnen Staaten deutlich und auch der Veränderungsbedarf im Bildungswesen kann hier, allerdings ‚nur‘ auf der Ebene eines Staates, dargestellt werden.

### 3.2.2 ‚Literacy‘-Orientierung vs. curriculare Orientierung

Ganz allgemein kann der Begriff ‚literacy‘ mit dem Begriff ‚Literalität‘ übersetzt werden. In der deutschen Literatur wird mit dem Begriff der ‚Literalität‘ häufig allerdings nur die Fähigkeit zu lesen und zu schreiben bezeichnet. Der vorliegende Studienbrief-Teil schließt sich einer weiter gefassten und in diesem Kapitel beschriebenen Definition an und verwendet im Folgenden den Begriff ‚literacy‘, ohne ihn zu übersetzen. Diese Variabilität des weit gefassten Begriffs der ‚literacy‘ wird auch dadurch deutlich, dass von ‚computer literacy‘, ‚science literacy‘ oder auch ‚health literacy‘ gesprochen werden kann (Grotluschen/Linde, 2007), sodass die in ‚literacy‘ beinhalteten Kompetenzen weit über das reine Schreiben und Lesen hinausgehen, diese Kompetenzen aber weiterhin wichtiger Bestandteil der ‚literacy‘ sind.

Nach einer solchen Auffassung „gehören die Beherrschung der Muttersprache in Wort und Schrift sowie ein hinreichend sicherer Umgang mit mathematischen Symbolen und Modellen zum Kernbestand kultureller Literalität“ (Köller, 2008). Dazu gehören „sprachliche Kompetenzen, die grundlegende Formen des kommunikativen Umgangs mit der Welt repräsentieren“ (Klieme/Baumert/Köller, 2000).

Aber auch eine ‚computer literacy‘ gehört zu den Kulturwerkzeugen der heutigen Welt. All diese Kompetenzen dienen in Kombination dazu, dass Menschen in ihrem Alltag, im Beruf und in der Öffentlichkeit selbstverantwort-

<sup>9</sup> In Vergleichsgruppe 1 befinden sich die Staaten der europäischen Union, die an IGLU teilgenommen haben.

<sup>10</sup> Vergleichsgruppe 2 beinhaltet alle Teilnehmerstaaten der OECD ohne die Türkei, dafür alle Staaten, die bis 2007 Mitglied der europäischen Union wurden.

<sup>11</sup> Vergleichsgruppe 3 klassifiziert sich nach den Ergebnissen in IGLU: Hier werden die Staaten aufgeführt, die beim Lesen mehr als eine halbe Standardabweichung unterhalb des internationalen Mittelwertes liegen.

lich handeln können (Klieme u.a., 2001). Defizite in diesen Bereichen der Grundbildung führen dazu, dass die Teilnahme an der allgemeinen gesellschaftlichen Entwicklung gefährdet ist und Perspektiven in den Bereichen ‚Beruf‘ und ‚Leben‘ schlechter sind (Köller, 2008).

Defizite

Bestimmte Bereiche der Grundbildung sind aber ‚wichtiger‘ als andere. In Bezug auf ‚scientific literacy‘ heißt das, dass die meisten Menschen mit einem naturwissenschaftlichen Alltagswissen auskommen, naturwissenschaftliches Spezialwissen dagegen nicht beherrscht werden muss. Hier unterscheidet man zwischen ‚scientific literacy‘, die nur von etwa 10 Prozent eines Jahrgangs erreicht wird, und ‚functional literacy‘, d.h. Alltagswissen, über das etwa 30 Prozent eines Jahrgangs verfügen (Baumert/Lehmann, 1997).

‚wichtigere‘  
Bereiche

### Vertiefung: Anpassung des ‚literacy‘-Konzepts in PISA (optional)

Für jeden Bereich der ‚Grundbildung‘ wird das ‚literacy‘-Konzept entsprechend angepasst (Kiper, 2003b). Als Beispiel werden die Definitionen von ‚scientific literacy‘, ‚reading literacy‘ und ‚mathematical literacy‘ dargestellt, wie sie in PISA verwendet werden.

#### Definition: ‚scientific literacy‘ in PISA

„Naturwissenschaftliches Wissen anzuwenden, naturwissenschaftliche Fragen zu erkennen und aus Belegen Schlussfolgerungen zu ziehen, um Entscheidungen zu verstehen und zu treffen, die die natürliche Welt und die durch menschliches Handeln an ihr vorgenommenen Veränderungen betreffen.“ (PISA-Konsortium, 2001)

#### Definition: ‚reading literacy‘ in PISA

„[...] die Fähigkeit, geschriebene Texte unterschiedlicher Art in ihren Aussagen, ihren Absichten und ihrer formalen Struktur zu verstehen und sie in einen größeren sinnstiftenden Zusammenhang einzuordnen, sowie in der Lage zu sein, Texte für verschiedene Zwecke sachgerecht zu nutzen.“ (PISA-Konsortium, 2001)

#### Definition: ‚mathematical literacy‘ in PISA

„[...] ein Verständnis der Rolle, die Mathematik in der sozialen, kulturellen und technischen Welt spielt, und die Fähigkeit, Sachverhalte unter mathematischen Gesichtspunkten angemessen zu beurteilen: [...] auch die Fähigkeit [...] Mathematik aktiv zu nutzen, um Anforderungen des Alltags zu bewältigen.“ (PISA-Konsortium, 2001)

Innerhalb der ‚literacy‘-Orientierung kann man mehrere Stufen unterscheiden, die „von einem anschaulichen Verständnis von Alltagsphänomenen über ein sinnvolles Anwenden elementarer Modelle bis hin zur vollen Kommunikations- und Urteilsfähigkeit in diesen Gebieten reichen“ (Klieme u.a., 2001).

Vor Beginn einer Studie des Systemmonitorings können sich die Administratoren beziehungsweise die Auftraggeber entscheiden, ob sie die im Curriculum beschriebenen Kompetenzen oder aber ‚literacy‘ als Grundlage untersuchen wollen, sodass sich eine ‚literacy‘-Orientierung oder eine curriculare Orientierung ergibt. Bei beiden Orientierungen handelt es sich um Formen der kriterialen Bezugsnorm (vgl. Kapitel 3.2.1.1).

Als Beispiel soll TIMSS dienen, da hier beide Orientierungen in der Studie vertreten sind (Klieme u.a., 2001).

### Beispiel: curriculare und ‚literacy‘-Orientierung in TIMSS (1995)

#### Curriculare Orientierung:

TIMSS II und TIMSS III lehnten sich eng an die jeweiligen Lehrpläne an. Ziel war es hier, „zentrale Inhalte und Anforderungen des Fachunterrichts“ (Klieme u.a., 2001) zu untersuchen. Dies ließ sich vor allem dadurch gut ermöglichen, dass die Curricula international relativ übereinstimmend waren.

#### ‚literacy‘-Orientierung:

Die mathematische und naturwissenschaftliche Grundbildung in TIMSS III wurde nach dem ‚literacy‘-Konzept untersucht. Unter Grundbildung wurde hier „die Kenntnis zentraler Konzepte und Arbeitsprinzipien dieser Fächer“, aber auch „die Fähigkeit, dieses Wissen in alltäglichen Zusammenhängen zu nutzen und zu kommunizieren“ verstanden. Hier wird besonders deutlich, dass ‚literacy‘ „eine Verknüpfung von fachsystematischem Verständnis und Anwendungsorientierung“ beschreibt.

Dabei orientierte sich der TIMSS-Grundbildungstest an folgenden Merkmalen des ‚literacy‘-Konzepts (Klime/Baumert/Köller, 2000):

- Betonung zentraler theoretischer Konzepte
- Einschränkung der stofflichen Breite zu Gunsten der Möglichkeit, in einzelnen Gebieten tieferes Verständnis zu erreichen
- Verstärkung fachübergreifender und fächerverbindender Ansätze
- Betonung des selbstständigen mathematischen und naturwissenschaftlichen Handelns und Kommunizierens

Die Vermischung der beiden Orientierungen in TIMSS wurde allerdings stets kritisch diskutiert. Im Gegensatz dazu steht PISA: Hier wird die curriculare Orientierung in den Hintergrund gedrängt und stattdessen das ‚literacy‘-Konzept verfolgt (PISA-Konsortium, 2002). Das, was in PISA unter ‚Bildung‘ verstanden wird, „orientiert sich damit zunächst an der Funktion von Kompetenzen im kulturellen und gesellschaftlichen Zusammenhang“. Bei der Ausarbeitung der Grundkonzeption von PISA wurde hinterfragt, welches Konzeptwissen („Wissen, dass“) und welches Prozesswissen („Wissen, wie“) für die Teilhabe am gesellschaftlichen Leben wichtig ist und in welchen Situationen welches Wissen angewendet werden soll (PISA-Konsortium Deutschland, 2007).

#### Vertiefung: Konzept- und Prozesswissen (optional)

<b>Konzeptwissen (Wissen, dass)</b>		
Definition: Wissen über die Wechselbeziehungen zwischen den einzelnen Elementen des Basiswissens innerhalb eines größeren Zusammenhangs, das ein gemeinsames Funktionieren sichert		
Kenntnis der Klassifikationen und Kategorien (z. B. die verschiedenen geologischen Zeitperioden)	Kenntnis der Prinzipien und Verallgemeinerungen (z. B. Theoreme und Gesetze)	Kenntnis der Theorien, Modelle und Strukturen (z. B. Evolutionstheorie)

(nach: ThILLM – Thüringer Institut für Lehrerfortbildung, Lehrplanentwicklung und Medien; Gehirngerechtes Klassenzimmer – „Handreichungen für die Unterrichtspraxis“; ThILLM-Heft 126)

<b>Prozesswissen (Wissen, wie)</b>		
Definition: Wissen darüber, wie man etwas tut; Wissen über Methoden des Nachforschens; Kenntnis über Anwendungskriterien für Fähigkeiten, Algorithmen, Techniken und Methoden		
Kenntnis fachspezifischer Fähigkeiten und Algorithmen (z. B. zur Lösung einer quadratischen Gleichung)	Kenntnis fachspezifischer Techniken und Methoden (z. B. über die Interpretation eines literarischen Textes)	Kenntnis der Kriterien zur Anwendung bestimmter Verfahrensweisen (z. B. welche Methoden zu benutzen sind, um Informationen eines Textes zu visualisieren)

(nach: ThILLM – Thüringer Institut für Lehrerfortbildung, Lehrplanentwicklung und Medien; Gehirngerechtes Klassenzimmer – „Handreichungen für die Unterrichtspraxis“; ThILLM-Heft 126)

Auch in IGLU orientiert sich das Konzept der Lesekompetenz an ‚literacy‘, sodass unter ‚reading literacy‘ hier die Kompetenz gemeint ist, „Lesen in unterschiedlichen, für die Lebensbewältigung praktisch bedeutsamen Verwendungssituationen einsetzen zu können“ (Bos u.a., 2003). Somit ergibt sich die theoretische Struktur der Lesekompetenz wie in Abbildung 3 dargestellt:

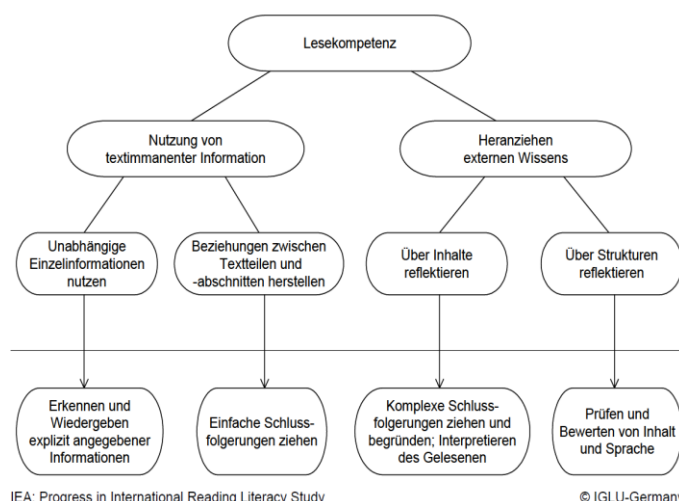


Abbildung 3: Theoretische Struktur der Lesekompetenz in IGLU

Die Abbildung zeigt, dass die Lesekompetenz bei IGLU in zwei theoretische Dimensionen eingeteilt wird. Dies ist einerseits die Verstehensleistung von textimmanenten, andererseits die Verstehensleistung von wissensbasierten Informationen durch externes Wissen. Diese beiden Dimensionen werden dann in weitere Aspekte unterteilt, die sich durch die unterschiedlichen kognitiven Anforderungen ergeben (Bos u.a., 2003).

Hierbei wird deutlich, dass Lesen nicht nur für die reine Wissensaneignung sondern vielmehr für den alltäglichen Gebrauch benutzt werden soll. Zudem soll sichergestellt werden, dass auch nach Abschluss der Schule die Lesekompetenz so gestaltet ist, dass die Schülerinnen und Schüler auf weiteres Lernen vorbereitet sind (Bos u.a., 2003).

Ein Beispiel für eine Studie mit curriculärer Orientierung ist DESI (vgl. Kapitel 3.1.3.4). Bei DESI wurde vor Beginn der Testentwicklung eine detaillierte Lehrplananalyse durchgeführt, sodass die Aufgaben in Inhalt und im Anspruch eng an die Lehrpläne angelehnt sind (Klieme, 2006).

Möglich ist, dass sich durch die internationalen Studien des Systemmonitorings eine curriculare Vereinheitlichung in den OECD-Staaten entwickeln wird. In einem solchen Curriculum werden dann kulturelle und nationale Besonderheiten immer weniger bedeutsam. Das ‚literacy‘-Konzept kann bei der Entwicklung eines solchen Curriculums behilflich sein, da dadurch den Studien bereits ein internationales Curriculum zur Verfügung steht. Somit erfolgt „eine Verschiebung in Richtung Wertschätzung anwendungsorientierten Wissens und handlungsbezogener Kompetenzen“ (Kiper, 2003b). Was in diesem Zusammenhang genau unter ‚Kompetenzen‘ verstanden wird, wird im folgenden Kapitel erläutert.

Lesen für den Alltag

DESI

curriculare Vereinheitlichung

### 3.2.3 Kompetenz, Kompetenzniveaus, Kompetenzmodelle und Standards

Das Kapitel beginnt mit einer Definition des Begriffs ‚Kompetenz‘. Es folgen Ausführungen zu ‚Kompetenzmodellen‘ und ‚Kompetenzniveaus‘ und deren Verbindung mit Standards. Damit keine Doppelungen mit den beiden anderen Studienbrief-Teilen entstehen, wird darauf geachtet, dass alle Erläuterungen eine Verbindung zum Systemmonitoring aufweisen.

#### 3.2.3.1 Kompetenz

Im Studienbrief „Vergleichsarbeiten“ wird der Begriff ‚Kompetenz‘ eingehend definiert, daher soll an dieser Stelle darauf verwiesen werden und hier nur einige ergänzende Aspekte mit Bezug auf das Systemmonitoring präsentiert werden.

Die Aufgaben in den Studien des Systemmonitorings fragen kein reines Faktenwissen ab, sondern untersuchen die Möglichkeiten der Schülerinnen und Schüler, in bestimmten Kontexten Anforderungen zu erfüllen und Probleme zu lösen. Je nachdem wie die Schülerinnen und Schüler die Aufgaben beantworten, wird ihnen eine entsprechende Kompetenz zugesprochen (PISA-Konsortium, 2008).

Somit beschreibt ‚Kompetenz‘ die Fähigkeit, bestimmte Probleme und Anforderungen zu lösen. Das folgende Beispiel zeigt, wie eine Fremdsprachenkompetenz beschrieben werden kann, die diese Aspekte beachtet. Das zu erreichende Ziel beziehungsweise das zu lösende Problem besteht hier in der kommunikativen Handlungsfähigkeit.

**Beispiel: Fremdsprachenkompetenz**

(Bildungsziel: kommunikative Handlungsfähigkeit) drückt sich darin aus, ...

- ... wie gut man kommunikative Situationen bewältigt
- ... wie gut man Texte unterschiedlicher Art versteht
- ... dass man adressatengerecht Texte verfassen kann
- ... dass man grammatische Strukturen korrekt aufbauen und bei Bedarf korrigieren kann
- ... dass man sich in der Intention und Motivation offen und akzeptierend mit anderen Kulturen auseinander setzen kann (Klieme, 2007).

Kompetenzen  
abgrenzbar

Von besonderer Bedeutung ist es, dass Kompetenzen voneinander abgrenzbar sind. Das heißt, dass eindeutig bestimmt werden kann, ob eine Schülerin oder ein Schüler eine bestimmte Kompetenz besitzt oder nicht besitzt. Daher werden die Kompetenzen sehr konkret beschrieben, sodass sie leicht in entsprechende Aufgaben umgesetzt und getestet werden können (KMK, 2005).

Vorteile

Dass im Unterricht und in den Studien zum Systemmonitoring Kompetenzen untersucht werden, hat die folgenden Vorteile (KMK, 2005):

- Der Blick wird auf die Lernergebnisse der Schülerinnen und Schüler gelenkt.
- Das Lernen wird auf die Bewältigung von Anforderungen und nicht nur auf den Aufbau von - zunächst - ungenutztem Wissen ausgerichtet.
- Das Lernen wird als kumulativer Prozess organisiert.

**3.2.3.2 Kompetenzmodelle**

Die Aufgaben und Charakteristika von Kompetenzmodellen sind:

Aufgabe der  
Modelle

- zu beschreiben, welche Lernergebnisse von Schülerinnen und Schülern zu bestimmten Zeitpunkten in den einzelnen Fächern erwartet werden
- aufzuzeigen, auf welche Art und Weise Wissen und Können erreicht werden können
- dabei zu helfen, zwischen abstrakt formulierten Bildungszielen und Aufgabensammlungen zu vermitteln
- die Basis dafür zu legen, dass Bildungsziele operationalisiert (d.h. messbar gemacht werden) und diese durch empirische Tests überprüft werden können
- nicht nur bei der Herstellung von Testverfahren, sondern auch in der Unterrichtspraxis zu helfen
- Anhaltspunkte zu bieten, sodass sich der Unterricht an den „Lernprozessen und Lernergebnissen der Schülerinnen und Schüler im jeweiligen Lernbereich“ orientieren kann (Klieme, 2007).

Entwicklung  
der Modelle

Die Kompetenzmodelle werden von internationalen Expertengruppen auf Basis der „pädagogischen und fachdidaktischen Forschung“ entwickelt (Klieme, 2007). Eine Anforderung an Kompetenzmodelle ist, sowohl im oberen, mittleren, als auch im unteren Leistungsbereich auszudifferenzieren. Unter ‚Ausdifferenzierung‘ wird dabei verstanden, dass für das jeweilige Kompetenzniveau ganz genau beschrieben wird, was die Schülerinnen und Schüler können müssen, um auf diesem Niveau eingestuft zu werden.

Es müssen aber nicht nur testtheoretische, fachliche und fachdidaktische Kriterien beachtet werden, sondern die Kompetenzmodelle müssen auch bildungspolitischen Erwartungen und pädagogischen Erfordernissen entsprechen. Als Beispiel soll hier angeführt werden, wie diese Erwartungen und Erfordernisse für die Kompetenzmodelle der Bildungsstandard-Überprüfung formuliert werden (KMK, 2009):

- herausfordernde und zugleich angemessene Leistungserwartungen beschreiben, die der Leistungsstreuung innerhalb und zwischen den Bundesländern in angemessener Weise Rechnung tragen
- trotz der zu erwartenden unterschiedlich hohen Anteile von Schülerinnen und Schülern, die den Mindest- oder Regelanforderungen nicht entsprechen, für alle Länder ein ‚Leistungsminimum‘ beschreiben, das von allen Schülerinnen und Schülern mittelfristig erreicht wird
- vorhandene sowie auszubauende Leistungsressourcen verdeutlichen
- motivierende Leistungserwartungen formulieren, die Entwicklungsimpulse an den Schulen auslösen
- breite bildungspolitische Akzeptanz insbesondere bei den Lehrkräften erreichen
- in einer spannungsreichen Relation zu den internationalen Ergebnissen stehen.

Der weitere Bestandteil eines Kompetenzmodells, nämlich die Einteilung in Abstufungen, soll im nächsten Kapitel unter dem Begriff ‚Kompetenzniveau‘ betrachtet werden.

**3.2.3.3 Kompetenzniveaus**

Wie bereits beschrieben, ist es nicht nur wichtig zu untersuchen, aus welchen Teilen sich eine bestimmte Kompetenz in einem Kompetenzmodell zusammensetzt, sondern auch, wie diese Kompetenz abgestuft werden kann. Auf Grund dieser Kompetenzniveaus wird es ermöglicht, dass Testergebnisse kriterienorientiert interpretiert werden können (vgl. Kapitel 3.2.1.1).

Jedes Niveau ist durch „kognitive Prozesse und Handlungen von bestimmter Qualität“ spezifiziert. Dabei sind dies immer Fähigkeiten, die die Schülerinnen und Schüler nur auf einem bestimmten Niveau, nicht aber auf einem darüber liegenden Niveau beherrschen (Klieme, 2007).

Beispiele solcher Stufenmodelle gibt es bei TIMSS und PISA, aber auch bei der Normierung der Bildungsstandards. Dabei wird davon ausgegangen, dass sich für jede Stufe bestimmte Aufgaben erstellen lassen, sodass Schülerinnen und Schüler auf Kompetenzniveau I die Aufgaben von Niveau I mit sehr großer Wahrscheinlichkeit lösen können, die Aufgaben auf Niveau II dagegen nur mit einer sehr geringen Wahrscheinlichkeit. Wie die einzelnen Stufen in Aufgaben umgesetzt werden, wird in Kapitel 3.3 beschrieben.

Im Folgenden sollen die Kompetenzniveaus von PISA im Bereich Mathematik<sup>12</sup> dargestellt werden. Hier wird deutlich, dass auf der unteren Kompetenzstufe Schülerinnen und Schüler zugeordnet sind, die „ein arithmetisches Wissen besitzen, das sie abrufen und anwenden können“, auf der obersten Stufe dagegen „komplexe Modellierungen und mathematische Argumentationen“ geleistet werden (Klieme, 2007).

### **Beispiel: Kompetenzniveaus in PISA: Mathematik**

#### Kompetenzniveau I: Rechnen auf Grundschulniveau

Personen, die diesem Niveau zugeordnet werden, verfügen lediglich über arithmetisches und geometrisches Wissen auf Grundschulniveau. Sie können dieses Wissen abrufen und unmittelbar anwenden, wenn die Aufgabenstellung von vornherein eine bestimmte Standard-Mathematisierung nahe legt. Begriffliche Modellierungen sind nicht leistbar.

#### Kompetenzniveau II: Elementare Modellierungen

Auf diesem Niveau werden auch einfachste begriffliche Modellierungen vorgenommen, die in einem außermathematischen Kontext eingebettet sind. Personen auf diesem Kompetenzniveau können unter mehreren möglichen Lösungsansätzen den passenden finden, wenn durch Graphiken, Tabellen, Zeichnungen usw. eine Struktur vorgegeben ist, die das Modellieren erleichtert. Auch auf diesem Niveau sind allerdings nur die Wissensinhalte der Grundschulmathematik sicher verfügbar.

#### Kompetenzniveau III: Modellieren und begriffliches Verknüpfen auf dem Niveau der Sekundarstufe I

Mit diesem Niveau findet im Vergleich zu Niveau II in mehrfacher Hinsicht ein qualitativer Sprung statt. Schülerinnen und Schüler auf diesem Kompetenzniveau verfügen auch über einfache Wissensinhalte der Sekundarstufe I, also über den Standardstoff der Lehrpläne aller Schulformen. Sie können Konzepte aus unterschiedlichen mathematischen Bereichen verknüpfen und zur Lösung von Problemstellungen nutzen, wenn visuelle Darstellungen den Lösungsprozess unterstützen.

#### Kompetenzniveau IV: Umfangreiche Modellierungen auf der Basis anspruchsvoller Begriffe

Schülerinnen und Schüler auf diesem Kompetenzniveau bewältigen im technischen Bereich umfangreichere Verarbeitungsprozesse, können also eine Lösung über mehrere Zwischenergebnisse hinweg aufbauen. Auch offene Modellierungsaufgaben werden bewältigt, bei denen man unter vielfältigen Lösungswegen einen eigenen finden muss. Verstärkt können auch innermathematische begriffliche Zusammenhänge modelliert werden.

#### Kompetenzniveau V: Komplexe Modellierung und innermathematisches Argumentieren

Auf diesem letzten Niveau ist auch anspruchsvolles curriculares Wissen verfügbar. Die Schülerinnen und Schüler, die diesem Kompetenzniveau zugeordnet werden, können auch sehr offen formulierte Aufgaben bewältigen, bei denen ein Modell frei gewählt beziehungsweise selbst konstruiert werden muss. Begriffliche Modellierungsleistungen auf dieser höchsten Stufe umschließen häufig Begründungen und Beweise sowie das Reflektieren über den Modellierungsprozess selbst.

(nach: Klieme/Neubrand/Lüdtke, 2001)

<sup>12</sup> Entsprechende Stufeneinteilungen existieren für alle in PISA getesteten Kompetenzen.



### 3.2.3.4 Standards

Auf Grund der Kompetenzmodelle und der Kompetenzniveaus ist es möglich, Standards jahrgangs- und fachbezogen festzulegen. Standards werden verbal formuliert und stützen sich auf die beschriebenen Modelle und Stufen. So können Mindest- oder Regelstandards festgehalten werden.

#### Definition: Mindeststandards

Mindeststandards beschreiben die Kompetenzen, die alle Schülerinnen und Schüler als Minimum erreichen sollen. Mindeststandards können erst dann formuliert werden, wenn die Standards bereits einen „Prozess der Erfahrung“ durchlaufen haben. Wird dieser Prozess, in dem Aufgabenbeispiele validiert und Schwierigkeiten von Aufgaben getestet werden, nicht eingehalten, so kann es passieren, dass Schülerinnen und Schüler entweder über- oder unterfordert werden. Für die Mindeststandards sind Kompetenzniveaus eine unabdingbare Voraussetzung.

(Klieme, 2007; KMK, 2005)

#### Definition: Regelstandards

Regelstandards beschreiben die Fähigkeiten, die in der Regel von den Schülerinnen und Schülern erreicht werden sollen. Dabei werden die Fähigkeiten benannt, die von mindestens der Hälfte der Schülerinnen und Schüler erreicht werden sollen, also ein mittleres Anforderungsniveau. Die von der KMK formulierten Bildungsstandards sind beispielsweise Regelstandards. Sie basieren auf Einschätzungen von Experten aus Theorie und Praxis. Auch diese Standards müssen nach ihrer Einführung noch validiert werden.

(Klieme, 2007; KMK, 2005)

### 3.2.4 Weiterführende Literatur

Als Literatur zum ‚literacy‘-Konzept empfiehlt sich Klieme/Baumert/Köller 2000. Auch der Aufsatz von Hanna Kiper (2003b) „literacy versus Curriculum?“ gibt einen guten Einblick in das Konzept inklusive der Vor- und Nachteile. Zudem beschreiben die jeweiligen Berichte der einzelnen Studien (vgl. Kapitel 3.1.3.1 bis 3.1.3.4), ob und wie in ihnen das Konzept von ‚literacy‘ ein- und umgesetzt wird. Klieme (2007) beschreibt die Bereiche der Kompetenz im Rahmen der Standardentwicklung sehr ausführlich. Weitere Informationen über den Kompetenzbegriff in den jeweiligen Studien und die Einteilung in Kompetenzniveaus kann der Literatur zu den Studien entnommen werden (vgl. Kapitel 3.1).

### 3.2.5 Verständnis-Aufgaben und Diskussionspunkte

1. *Wer Leistungsvergleiche durchführen will, braucht Maßstäbe. Welcher Maßstab eignet sich für welche Problemstellung? Worin bestehen Vor- und Nachteile?*
2. *Konstruieren Sie einige Aufgaben in einem gewählten Fachbereich auf unterschiedlichen Kompetenzniveaus und dokumentieren Sie den Prozess.*
3. *‚Literacy‘-Orientierung und curriculare Orientierung sind unterschiedliche Konzepte bei der Konstruktion eines Untersuchungsdesigns. Welche möglichen Auswirkungen hat die Auswahl auf die Schul- und Unterrichtsentwicklung?*
4. *Es gibt sowohl in der Wissenschaft als auch in der Politik eine Diskussion, die statt der/ergänzend zu den Regelstandards die Definition von Mindeststandards fordert. Reflektieren Sie diese Position und ihre Konsequenzen für das Bildungswesen.*

### 3.3 Testkonstruktion

Kapitel 3.3 beschäftigt sich mit der Testkonstruktion. Zunächst wird der theoretische Hintergrund beleuchtet, es folgen Betrachtungen der Testgütekriterien, der Test-Verfahrensweisen, der Test-Aufgabenauswahl und der Zielpopulation. Das Kapitel endet mit einer Übersicht über die bisher durchgeführten und geplanten Erhebungen im Rahmen des Systemmonitorings.

In diesem Kapitel werden folgende Fragen beantwortet:

- *Welchen Anforderungen müssen die Aufgaben der Studien genügen? Und warum?*
- *Wie verläuft die Testdurchführung?*
- *Wie erfolgt die Auswertung beziehungsweise Analyse der Ergebnisse?*
- *Wie werden Aufgaben für die Studie ausgewählt und wie werden sie in den Kompetenzmodellen verortet?*
- *Wie werden die Schülerinnen und Schüler ausgesucht, die an der Testdurchführung teilnehmen?*
- *Welche Studien wurden bereits durchgeführt und welche sind in Planung?*

Auch wenn die Zugänge zu diesen Fragestellungen eher theoretischer Natur sind, dienen sie im Besonderen dem Verständnis der Systemmonitoring-Studien. Von herausragender Bedeutung ist die Betrachtung der „professionellen Qualitätsmaßstäbe“, um der Gefahr von Fehl- und Überinterpretationen zu begegnen (Klieme u.a., 2007).

Die in diesem Kapitel gegebenen Informationen sind wichtig, da so das Prinzip verstanden werden kann, das hinter den Systemmonitoring-Studien steht. Zudem ist eine Interpretation der Ergebnisse einfacher, wenn die Hintergründe bekannt sind, unter denen die erzielten Punktwerte der Schülerinnen und Schüler entstanden sind.

Im gesamten Kapitel sollen dabei ausschließlich die international durchgeführten Studien betrachtet werden.

#### 3.3.1 Testtheoretischer Hintergrund

Die wichtigste Voraussetzung für das Testen schulischer Leistungen ist, dass die Leistung sich durch direkt beobachtbares Verhalten festhalten lässt. Dies können sprachliche Äußerungen oder Aufgabenbearbeitungen sein.

Die Untersuchung muss so gestaltet werden, dass die unterschiedlichen Leistungspotenziale der Schülerinnen und Schüler dadurch erkennbar werden, dass sie sich ihrem Leistungsniveau entsprechend unterschiedlich verhalten. Diesen Leistungsniveaus werden bestimmte Skalenwerte zugeordnet. Für die Zwecke des Systemmonitorings wird versucht, ein Ordinal- oder Intervallskalenniveau (vgl. Erklärung unten) zu verwenden, da die Unterschiede zwischen den einzelnen Messwerten so möglichst aussagekräftig sind (Heller/Hany, 2001).

Leistung =  
beobachtbares  
Verhalten

#### Vertiefung: Ordinal- und Intervallskala (optional)

##### Definition: Ordinalskala

„Die relative Größe der den Objekten zugeordneten Zahlen reflektiert die Ausprägung des Merkmals, die die Objekte besitzen. Gleiche Differenzen zwischen den Zahlen implizieren nicht gleiche Differenzen in den Merkmalsausprägungen.“

Beispiele: Rangplätze im Sport, Schulnoten, Rangbildung hinsichtlich eines psychologischen Merkmals, Testrohwerte, Härte von Mineralien

(Diehl/Kohr, 2004)

##### Definition: Intervallskala

„Es existiert eine Maßeinheit, mit der die Objekte nicht nur zugeordnet werden können – man kann ihnen auch Zahlen so zuweisen, dass gleiche Differenzen zwischen den den Objekten zugeordneten Zahlen gleiche Differenzen in der Ausprägung des gemessenen Merkmals reflektieren. Der Nullpunkt ist willkürlich gesetzt und bedeutet nicht ‚Nichtvorhandensein‘ des Merkmals.“

Beispiele: Kalenderzeit, Celsius und Fahrenheit, Temperaturskalen, Intelligenztestwerte

(Diehl/Kohr, 2004)

Es kann ein Modell der probabilistischen Testtheorie, das sogenannte Rasch-Modell (vgl. Studienbrief-Teil Vergleichsarbeiten), angewendet werden, um Fähigkeiten der Schülerinnen und Schüler und Schwierigkeiten der Items auf der gleichen Skala abzubilden. Übersteigt die Fähigkeit eines Schülers die Schwierigkeit eines Items,

Probabilistische  
Testtheorie /  
Rasch-Modell



kann man davon ausgehen, dass dieser Schüler dieses Item mit größerer Wahrscheinlichkeit richtig beantwortet, als dass er es falsch beantwortet.

Im Bereich der Kompetenzdiagnostik wird die Erfassung von Kompetenzen mit Hilfe der probabilistischen Testtheorie auch als ‚Proficiency Scaling‘ bezeichnet. Die Testaufgaben werden dazu von Expertinnen und Experten des jeweiligen Inhaltsbereichs bezüglich ihrer Anforderungsmerkmale analysiert (vgl. Studienbrief-Teil Vergleichsarbeiten). Aufgaben-Cluster mit ähnlichen Anforderungsmerkmalen und ähnlicher Schwierigkeit werden im Anschluss als „inhaltlich definierte Stufen der Kompetenz“ (Bos u.a., 2003) beschrieben. Aufgabenbeispiele ermöglichen einen plastischen Eindruck der Anforderungen, denen Schülerinnen und Schüler auf dem jeweiligen Niveau gerecht werden (vgl. für eine detaillierte Beschreibung Studienbrief-Teil Vergleichsarbeiten).

### 3.3.2 Testgütekriterien und Testanforderungen

Für Informationen zur klassischen Testtheorie sowie eine ausführliche Beschreibung der drei grundlegenden Testgütekriterien Objektivität, Reliabilität und Validität sei auf die Ausführungen in Studienbrief-Teil ‚Individuelle Diagnostik‘ verwiesen. In diesem Kapitel sollen die Testgütekriterien nur in aller Kürze und jeweils mit einem Bezug zu den Studien des Systemmonitorings dargestellt werden. Dabei muss beachtet werden, dass die Bewertung eines Testes anhand der Testgütekriterien immer verbunden sein muss mit dem Blick auf die Schlussfolgerungen, die daraus gezogen werden sollen sowie deren Art und deren Nutzung (Klieme u.a., 2007).

Besonders für die internationalen Studien des Systemmonitorings ist es bedeutsam, dass Untersuchungsgegenstände ausgewählt werden, die Auskunft über die Leistungsfähigkeit der jeweiligen Bildungssysteme geben. Außerdem müssen Voraussetzungen, Durchführung und Auswertung der Tests für alle Teilnehmerstaaten gleich sein. Die Betrachtung der Validität soll sicherstellen, dass genau das gemessen wird, was auch gemessen werden soll (Heller/Hany, 2001). Dazu werden die Studien im Rahmen des Systemmonitorings sorgfältig konzipiert und die Aufgaben gewissenhaft ausgesucht. Zudem durchlaufen die Testaufgaben eine Erprobungsphase.

Um dem Kriterium der Objektivität gerecht zu werden, erhalten bei Studien des Bildungsmonitorings alle Testleiter von den Organisationen und Instituten, die die Studie konzipiert haben, genaue Anweisungen zur Durchführung der Studie. Darin ist genau beschrieben, wie die Aufgaben verteilt werden müssen, welche Anweisungen gegeben werden dürfen und welche Hilfestellungen und Hilfsmittel erlaubt sind. Die Einhaltung der Anweisungen wird von unangemeldeten Beobachtern überwacht.

Auch die Testauswertung muss konsistent durchgeführt werden, sodass gleiche Antworten zu einer gleichen Interpretation und Punktevergabe führen. Bei den großen Studien des Systemmonitorings wird die Testauswertung nicht von den Lehrerinnen und Lehrern, sondern von den Organisationen, die den Test entwickelt haben, durchgeführt. Hier werden Personen extra geschult um Ergebnisse zu erreichen, die dem Testgütekriterium der Objektivität entsprechen (Heller/Hany, 2001).

Die Reliabilität fordert, dass die Messungen präzise und unabhängig von Ort oder Zeit der Testdurchführung sind. Theoretisch müssen sich die Tests wiederholen lassen und das gleiche Ergebnis zeigen. Auch aus diesem Grund müssen die Anweisungen an die durchführenden Personen möglichst genau sein. Da es aber unmöglich ist, Messfehler zu vermeiden, ist es möglich und sinnvoll abzuschätzen, wie groß der Anteil der Messfehler sein kann (vgl. Studienbrief Teil 1).

### 3.3.3 Test-Verfahrensweisen

Die meisten Systemmonitoringstudien finden an zwei Tagen hintereinander statt, da die Ergänzungsstudien meist einen Tag nach den internationalen Erhebungen stattfinden. Für die internationale Durchführung wird ein Zeitfenster festgelegt, in dem die Durchführung stattfinden muss. Dieses Zeitfenster betrug beispielsweise bei PISA 2006 einen Monat (PISA-Konsortium, 2007). Speziell geschulte Testleiterinnen und Testleiter lesen den Schülerinnen und Schülern die Anweisungen vor und beaufsichtigen die Durchführung des Tests. Alle Aktivitäten der Testleiterinnen und Testleiter sind vorgegeben und müssen genau befolgt werden. Dies reicht von der Begrüßung über das Verteilen der Testhefte bis zum Erklären und Einsammeln der Testhefte (Bos u.a., 2003).

Das Rasch-Modell (vgl. Kapitel 3.3.1) ermöglicht es, einen an Items umfangreichen Test in zahlreiche Untertests aufzuteilen, welche die gleiche Kompetenz messen und von den Schülerinnen und Schülern zeitökonomischer bearbeitet werden können. Die Untertests werden im Rahmen der Testauswertung wieder zusammengefügt; die jeweiligen Kompetenz-(Fähigkeits-)Schätzungen auf Basis unterschiedlicher Testhefte ermöglichen es dennoch, alle Schülerinnen und Schüler auf der gleichen Skala abzubilden. Dieses Verfahren mit ‚rotierenden Testheften‘ entlastet die Schülerinnen und Schüler, da sie nur eine begrenzte Anzahl von Aufgaben bearbeiten müssen, ohne dass Nachteile für die Studie entstehen, da diese weiterhin eine hohe Repräsentativität erhalten kann. Zudem wird das Abschreiben der Schülerinnen und Schüler beinahe unmöglich gemacht, da sie unterschiedliche Aufgaben bearbeiten (Arnold, 2001).

Dieses Vorgehen nennt sich ‚Multi-Matrix-Design‘ oder ‚Multi(ple)-Matrix Sampling‘ und hat sich in der empirischen Bildungsforschung als Forschungsdesign etabliert. Die Teil-Tests oder Testhefte (bei IGLU 2001 gab es beispielsweise 12 verschiedene) ermöglichen es, dass die Schülerinnen und Schüler unterschiedliche Aufgaben be-

Proficiency  
Scaling

Validität

Objektivität

Testauswertung

Reliabilität

Zeitfenster

Teil-Tests /  
rotierende  
TesthefteMulti-  
Matrix-

arbeiten (Bos u.a., 2003). Durch dieses Vorgehen ist es möglich, dass die Schülerinnen und Schüler, wie beispielsweise in PISA 2006, nur 120 Minuten an den Aufgaben arbeiteten, sie jedoch als Gemeinschaft Aufgaben bearbeiteten, die 390 Minuten in Anspruch genommen hätten. Die Kompetenzen der Schülerinnen und Schüler konnten auf diese Art und Weise wesentlich umfangreicher untersucht werden. Allerdings verlangt ein solches Vorgehen auch eine komplexe statistische Auswertung (PISA-Konsortium, 2007).

### Vertiefung: Beispiel für Umgang mit rotierenden Testheften/Multi-Matrix-Design (optional)

Als Beispiel soll die Vorgehensweise bei PISA 2006 beschrieben werden. Hier erhielten die Schülerinnen und Schüler 13 unterschiedliche Testhefte. Dabei waren in jedem Testheft vier Aufgabengruppen vorhanden, die auch in drei weiteren Testheften vorkamen, sodass sich die Schwierigkeitsgrade der Aufgaben in den Aufgabengruppen unabhängig von den Testheften miteinander in Beziehung setzen ließen. Daraus folgt, dass in PISA 2006 jeder nur vier Dreizehntel der Aufgaben bearbeitete. Da sich die durchschnittlichen Schwierigkeiten der einzelnen Testhefte zwar ähnelten, aber nicht übereinstimmen, kann nicht die Anzahl der richtigen Lösungen als Maß für die Kompetenz der Schülerinnen und Schüler herangezogen werden. Daher wird auf Testmodelle der ‚Item-Response-Theorie‘ (IRT; vgl. Studienbrief-Teil Vergleichsarbeiten) zurückgegriffen, um die einzelnen Kompetenzwerte zu schätzen. Dabei geht man wie folgt vor:

„Man postuliert eine nicht direkt beobachtbare (latente) kontinuierliche Personenvariable, die das Antwortverhalten der Schülerinnen und Schüler erklärt. Zum Beispiel kann man annehmen, dass die naturwissenschaftliche Kompetenz eine latente Variable darstellt, mit der die Antworten auf die Aufgaben des PISA-Naturwissenschaftstests erklärt werden können“ (PISA-Konsortium, 2007).

Formalisiert werden kann dies durch ein Modell, in PISA ist dies das Rasch-Modell. Der größte Vorteil eines IRT-Modells im Vergleich zur klassischen Testtheorie (vgl. Studienbrief-Teil Individuelle Diagnostik) liegt darin, dass die Leistungen der Schülerinnen und Schüler auf einer gemeinsamen Skala abgebildet werden können, obwohl sie unterschiedliche Aufgaben bearbeitet haben (PISA-Konsortium, 2002). Weitere Eigenschaften eines solchen Skalierungsverfahrens, besonders für Studien des Systemmonitorings, sind zusammengefasst:

- Es ermöglicht, dass die inhaltliche Passung von Testaufgaben für die theoretisch bestimmte Fähigkeitsdimension empirisch geprüft werden kann und ungeeignete Aufgaben aussortiert werden können.
- Es ermöglicht, die Fähigkeit einer Person zuverlässig zu schätzen, auch wenn nur eine Untermenge der insgesamt verfügbaren Aufgaben bearbeitet wurde (s. oben).
- Es ermöglicht, Teilnehmergruppen mit unterschiedlicher Fähigkeit maßgeschneiderte Testversionen zu geben, da die Schwierigkeiten der bearbeiteten Aufgaben bei der Schätzung der Fähigkeit von Personen berücksichtigt werden.
- Es ermöglicht, die Schwierigkeit von Testaufgaben und die Fähigkeit von Personen auf demselben Maßstab abzubilden.

Allerdings kann das (dichotome) Rasch-Modell nur für Aufgaben verwendet werden, bei denen es nur richtige oder falsche Antworten gibt. Gibt es Aufgaben, bei denen die Schülerinnen und Schüler mehr als einen Punkt erlangen können, etwa weil für eine Aufgabe auch Teillösungen bewertet werden, muss ein anderes Modell gewählt werden. In diesem Fall kann das ‚Partial-Credit-Modell‘ (ordinales Rasch-Modell) für ordinale Items verwendet werden (PISA-Konsortium, 2007). Nähere Informationen dazu können der entsprechenden Literatur entnommen werden.

Die Untersuchungen bei IGLU dauerten an beiden Tagen etwa zwei Stunden (Bos u.a., 2003). Dies ist auch für die anderen Studien ein ungefährender Richtwert, es sollte darauf geachtet werden, dass die Bearbeitungsdauer altersangemessen ist. Wichtig ist es, dass, gerade bei jüngeren Schülerinnen und Schülern, abwechslungsreiche Aufgaben angeboten werden und die Möglichkeit besteht, eine Aufgabe zu überspringen und an einer anderen Stelle weiterzumachen (Bos u.a., 2003).

Zum Test gehören nicht nur die jeweiligen Aufgaben, sondern auch Fragebögen, die zusätzliche und ergänzende Informationen liefern können, um Leistungsunterschiede zu erklären. Diese werden von den Schülerinnen und Schülern im Rahmen der Testdurchführung ausgefüllt, die Eltern, Lehrerinnen und Lehrer und die Schulleitung füllen die Bögen ebenfalls während der zwei Testtage aus. Alle Unterlagen werden gemeinsam zur durchführenden Organisation zurückgeschickt (Bos u.a., 2003).

### 3.3.4 Testaufgabenwahl

Die Tests unterscheiden sich darin, welche Kompetenzen die Studie untersucht. Vor Beginn der Studie werden die zu untersuchenden Inhalte festgelegt, dabei kann ein bestimmtes Thema besonders ausführlich oder nur sehr knapp in die Planung eingehen. Bei Studien, die der ‚literacy‘-Orientierung folgen, orientieren sich die Inhalte

nicht an Lehrplänen. Bei PISA weichen beispielsweise die Aufgaben davon ab, was die Schülerinnen und Schüler aus dem Unterricht kennen, und zwar sowohl in Bezug auf die Aufgabenstellung und die damit verbundenen Anforderungen als auch auf die Themen und Inhaltsbereiche. So ist es nicht immer möglich, die Aufgaben aus den Studien direkt einem Schulfach zuzuordnen, da beispielsweise die Lesekompetenz als fächerübergreifende Kompetenz auch in den anderen Fächern involviert ist (Deutsches PISA-Konsortium, 2003). Die Studien versuchen, die Aufgaben möglichst nah am Alltag der Schülerinnen und Schüler anzusiedeln und trotzdem die fachlichen Kenntnisse zu überprüfen (Baumert/Bos/Watermann, 1999). Eine Veröffentlichung der Aufgaben nach Abschluss der Studien kann nur in sehr begrenztem Umfang stattfinden, da die Aufgaben aus methodischen Gründen für die nachfolgenden Erhebungswellen und einen längsschnittlichen Vergleich benötigt werden (van Ackeren, 2006).

#### Aufgabentypen

Die Aufgabentypen unterscheiden sich danach, welche Kompetenzen der Schülerinnen und Schüler getestet werden sollen. Auf Grund einer größeren Auswertungsobjektivität werden meist geschlossene Aufgaben ausgewählt, die nur eine richtige Antwort haben. Die bekanntesten und auch meist verwendeten Aufgaben sind dabei ‚Multiple-Choice-Aufgaben‘ (Mehrfachwahlaufgaben). Aber auch offene Aufgabenformate können gestellt werden, wie beispielsweise in IGLU (Bos u.a., 2003). Zu diesen offenen Aufgabenformaten gehören Aufgaben, die Kurzantworten verlangen (ein einziges Wort, mehrere Wörter, Zeichnungen). Meist werden die einzelnen Aufgaben in Gruppen zusammengefasst, wobei sich jeweils eine Gruppe der Aufgaben mit der Beschreibung einer bestimmten alltagsnahen Situation beschäftigt (PISA-Konsortium, 2002).

#### Rückmeldung

#### Aufgaben-entwicklung

Im Rahmen der Aufgabenentwicklung werden zunächst Aufgaben gesammelt, die für den Zweck der Studie theoretisch einsetzbar wären. Hierbei kann einerseits auf Aufgaben zurückgegriffen werden, die bereits in anderen Studien verwendet wurden, andererseits werden internationale Experten dazu aufgerufen, geeignete Aufgaben einzusenden. Beispielsweise für TIMSS wurde eine Aufgaben-Datenbank für die Aufgaben angelegt. Anschließend werden die Aufgaben von Experten begutachtet und die ausgewählten Aufgaben ausprobiert (Baumert/Lehmann, 1997). Zu beachten ist allerdings, dass TIMSS eine curriculare Orientierung verfolgt und daher die curriculare Validität durch diese Überprüfung sichergestellt werden sollte. Bei der Überprüfung der curricularen Validität der TIMSS-Aufgaben wurde auf die folgenden Kriterien geachtet (Baumert/Bos/Lehmann, 2000):

- Lehrplanvalidität des in einer Aufgabe repräsentierten Stoffes
- Vertrautheit mit der spezifischen Einbettung und Präsentation des Stoffes
- fachliche Qualität der Aufgabe unabhängig von ihrer curricularen Gültigkeit
- vermutete Lösungswahrscheinlichkeit der Aufgabe

#### Umgang mit unbekannten Aspekten

Eine solche Überprüfung der Angemessenheit der Aufgaben muss für Studien mit ‚literacy‘-Orientierung entsprechend angepasst werden. Wichtig ist, dass keine unbekannten Aspekte in den Aufgaben enthalten sind. Dies ist vor allem von großer Bedeutung, wenn die Schülerinnen und Schüler noch recht jung sind, da sie ansonsten überfordert sind und die Testdurchführung abbrechen könnten<sup>13</sup> (Bos u.a., 2008).

#### Erprobung der Aufgaben

In Pre-Tests<sup>14</sup> vor Beginn der eigentlichen Untersuchung werden die Aufgaben dann auf ihre Tauglichkeit untersucht.

### Vertiefung: Definition von Kompetenzniveaus (optional)

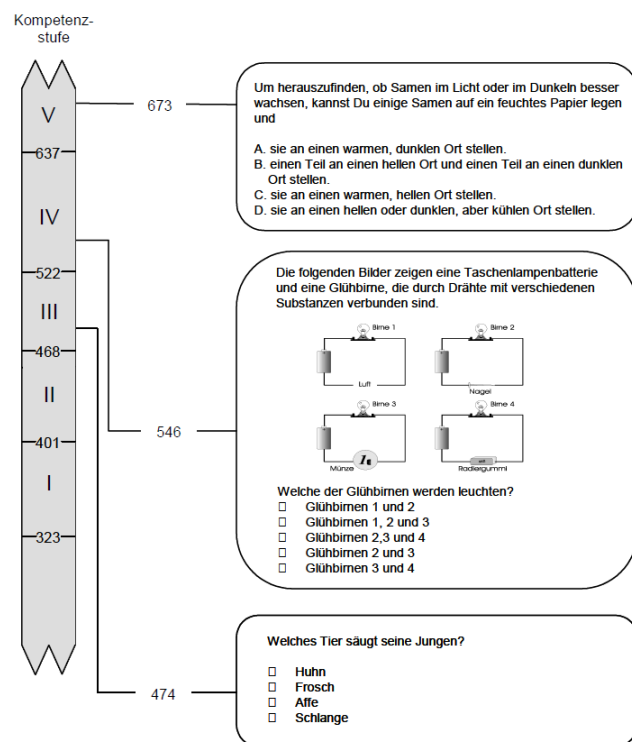
Durch die Nutzung der probabilistischen Testtheorie und der IRT-Skalierung ist es möglich, unterschiedliche Kompetenzniveaus zu definieren. In TIMSS III (1995) wurde beispielsweise so verfahren, dass zunächst Fixpunkte auf der Skala ausgewählt wurden. Um anschließend einen dieser Punkte als ‚Kompetenzniveau‘ zu definieren, wurden die Aufgaben begutachtet, die die Personen mit diesem Punktwert erreicht haben, mit hinreichender Sicherheit (65 Prozent) lösen konnten, nicht aber von Personen mit Punktwerten auf dem niedrigeren Niveau. Die für diese Aufgabe charakteristischen Anforderungen beschreiben dann das Kompetenzniveau (Baumert/Bos/Watermann, 1999).

Zudem ist es durch die Skalierung möglich, jeder Aufgabe einen Schwierigkeitswert zuzuordnen und damit einem Kompetenzniveau. Als Beispiel soll hier die Verteilung von drei Aufgaben auf die Kompetenzniveaus aus IGLU-E 2001 dienen.

<sup>13</sup> Eine solche Beurteilung, ob ein Aspekt bekannt oder unbekannt ist, kann von Experten durchgeführt werden, die beispielsweise mit Hilfe von Schulbüchern (bei curriculärer Orientierung) eine solche Einschätzung treffen. Dabei kann es sich um unbekannte graphische Darstellungsformen, mathematische Begriffe, Sachverhalte, Verfahren und Bezeichnungen handeln (Bos u.a., 2008).

<sup>14</sup> In Pre-Tests werden die geplanten Erhebungsinstrumente erprobt, die Ergebnisse fließen nicht mit in die Studie ein. Ziel ist es, dass für die eigentliche Untersuchung Fehlerquellen im Vorfeld erkannt und ausgeräumt werden können.

#### Definition von Kompetenzniveaus



IEA: Progress in International Reading Literacy Study

© IGLU-Germany

Abbildung 4: Verortung der Aufgaben nach Kompetenzniveaus bei IGLU

Das Beispiel macht deutlich, wo auf der Skala die Aufgaben ihrer Schwierigkeit nach anzuordnen sind.

Erst durch die Verknüpfung eines Niveaus mit einer Beschreibung der zu vollziehenden Aufgaben und Operationen kann der Maßstab des Niveaus verankert werden und eine Aussagekraft erhalten (Baumert/Lehmann, 1997). Eine solche Beschreibung wurde bereits in Kapitel 3.2.3.3 gegeben. Es handelt sich hier um eine kriteriale Bezugsnorm (vgl. Kapitel 3.2.1.1).

Verankerung  
des Maßstabs

Alle Schülerinnen und Schüler werden ihrem Ergebnis nach in eines der Kompetenzniveaus eingeteilt. Dadurch ist es möglich, den numerischen Wert des Ergebnisses inhaltlich zu verankern und ihm damit eine Aussagekraft zu verleihen (Baumert/Bos/Watermann, 1999).

Kompetenzniveau	Beschreibung	Skalenbereich der Fähigkeit
I	Gesuchte Wörter in einem Text erkennen	375-450
II	Angegebene Sachverhalte aus einer Textpassage erschließen	451-525
III	Implizit im Text enthaltene Sachverhalte aufgrund des Kontextes erschließen	526-600
IV	Mehrere Textpassagen sinnvoll miteinander in Beziehung setzen	> 600

Tabelle 8: Beschreibung der Kompetenzniveaus beim Leseverständnis und ihre Skalenwerte in IGLU (nach Bos u.a., 2003)

Häufig werden vier Kompetenzniveaus definiert. Das folgende Beispiel stammt aus IGLU und zeigt die vier Kompetenzniveaus, die Beschreibung der auf dieser Stufe erreichten Kompetenzen und den Skalenbereich, innerhalb dessen die Schülerinnen und Schüler liegen müssen, um einem bestimmten Kompetenzniveau zugeteilt zu werden. Entsprechende Beschreibungen werden in allen Studien des Systemmonitorings erstellt.

Kompetenz-  
niveaus

Wichtig ist dabei, dass die Ergebnisse der Skalierung in den Studien durch eine Transformation so standardisiert werden, dass alle Studien einen Mittelwert von 500 Punkten und eine Standardabweichung von 100 Punkten aufweisen.<sup>15</sup>

<sup>15</sup> Bei einer wiederholten Durchführung der Studie mit denselben Test-Items wird dann allerdings der tatsächlich errechnete Wert als Standard verwendet.

**Definition: Standardabweichung**

Die Standardabweichung kann als „Maß der durchschnittlichen Abweichung vom Mittelwert“ definiert werden. Aus ihr kann abgelesen werden, wie weit sich die Werte um den Mittelwert herum verteilen. Ein hoher Wert sagt aus, dass die Werte weit um den Mittelwert herum verteilt sind, ein kleiner Wert, dass die Werte nah am Mittelwert liegen. Zudem sagt die Standardabweichung auch etwas über die Reichweite der Wertverteilung aus (Albert/Koster, 2002).

Problem: Übersetzung der Aufgaben

Bei internationalen Untersuchungen muss darauf geachtet werden, dass alle Aufgaben in alle Sprachen gleich übersetzt werden. Darum werden beispielsweise bei IGLU die Aufgaben und auch die Fragebögen vom Englischen in die jeweilige Unterrichtssprache und wieder zurück übersetzt. Es wird kontrolliert, ob die Aufgaben dann noch immer mit der Ausgangsaufgabe übereinstimmen. Zudem werden die Aufgaben von Experten der Bundesländer auf ihre curriculare (inhaltliche) Validität begutachtet. Außerdem wird begutachtet, ob die Aufgaben angemessen in den Bereichen Sprache, Inhalt, Bearbeitungszeit und Vertrautheit mit der Textsorte (bei IGLU: Lesekompetenz) sind (Bos u.a., 2003).

Fragebögen

Innerhalb der Untersuchung werden aber nicht nur fachliche Inhalte untersucht, sondern auch weitere Faktoren, die die Leistungen der Schülerinnen und Schüler näher erklären sollen (Determinanten von Schulleistungen, vgl. Helmke & Schrader (1998), Studienbrief-Teil Vergleichsarbeiten). Vor allem internationale Unterschiede können so genauer betrachtet werden. Dazu werden Fragebögen konzipiert, die einerseits von den Schülerinnen und Schülern und ihren Eltern, andererseits aber auch von den Lehrerinnen und Lehrern oder den Schulleitungen ausfüllen werden. Welche Bereiche in diesen Fragebögen abgefragt werden, soll hier am Beispiel von IGLU dargestellt werden (Bos u.a., 2003):

- Familiärer Kontext
  - leseunterstützende Aktivitäten der Eltern
  - Sprache in der Familie
  - familiäre Ressourcen
  - Beziehungen zwischen Elternhaus und Schule
  - außerschulische Leseaktivitäten der Kinder
- Schulischer Kontext
  - Schulstandort und schulische Ressourcen
  - Lehrerbildung
  - Lernumgebung und Unterrichtsorganisation in der Klasse
  - Unterrichtsmethoden und -maßnahmen
  - Unterrichtsmaterial und -medien
- Nationaler und kommunaler Kontext
  - demographische Besonderheiten
  - Führung und Organisation des Bildungssystems
  - Lehrplanbesonderheiten und -schwerpunkte

Sowohl die Informationen aus den Fragebögen als auch aus den Testaufgaben werden streng vertraulich behandelt, alle Aspekte des Datenschutzes werden berücksichtigt. Die so erlangten Informationen dürfen nur zu wissenschaftlichen Zwecken verwendet werden (van Ackeren, 2006).

**3.3.5 Bezugsgruppen**

Damit verschiedene Gruppen von Schülerinnen und Schülern international verglichen werden können, müssen nicht nur die Inhalte, sondern auch die Stichproben der Populationen vergleichbar sein.

Ausmaß der Selektivität beachten

In den großangelegten internationalen Studien des Systemmonitorings ist es nicht möglich, die Grundgesamtheit der Schülerinnen und Schüler zu testen. Um Vergleiche zwischen Staaten oder Bundesländern ziehen zu können, muss das „Ausmaß der Selektivität der Stichproben, die gezogen und deren Kennwerte in den Vergleich genommen werden, abschätzbar“ gemacht werden. Studien der Sekundarstufen müssen beispielsweise beachten, dass in einigen Staaten verschiedene Schulformen existieren. Im Vergleich zu Staaten mit integrativen Systemen muss bei solchen mit einem gegliederten Schulsystem „diese Struktur durch eine so genannte Stratifizierung (d.h. schulformbezogene Aufteilung der Stichprobe und anschließende Gewichtung der Kennwerte) sehr genau nachgebildet werden“ (Arnold, 2001). Ebenfalls ist es problematisch, eine bestimmte Klassenstufe als Bestandteil der Population zu beschreiben, da das Alter der Schülerinnen und Schüler in einer bestimmten Klassenstufe international variiert. Daher wird vor Beginn der Untersuchung definiert, wer genau getestet werden soll. Eine Möglichkeit ist es, das Alter der teilnehmenden Schülerinnen und Schüler zu definieren (Beispiel: 15-jährige, PISA) oder



aber eine Klassenstufe, die dann näher beschrieben werden muss (Beispiel: Die Klassenstufe, in der die meisten Neunjährigen sind, IGLU).

Ein weiteres Problem entsteht, wenn die zu untersuchenden Schulsysteme unterschiedliche Quoten an Sonderschulen aufweisen und diese nicht mit in die Untersuchung einbezogen werden (Arnold, 2001). Aus diesem Grund werden in den internationalen Studien Quoten festgelegt, nach denen bestimmte Schülergruppen von der Untersuchung ausgeschlossen werden dürfen. Bei IGLU betrug diese Quote beispielsweise fünf Prozent der Population, ansonsten konnte das Land nicht in der internationalen Berichterstattung analysiert werden<sup>16</sup> (Bos u.a., 2003). Ausgeschlossen werden können auch Schülerinnen und Schüler, die die Testsprache nicht als Erstsprache sprechen und erst weniger als ein Jahr in dieser Sprache unterrichtet werden (PISA-Konsortium, 2007).

Umgang mit  
Sonderschulen

Umgang mit  
anderen  
Erstsprachen

Generell ist es wichtig, dass bei internationalen Untersuchungen die Stichprobe groß genug ist. Dies resultiert vor allem aus dem komplexen Stichproben-Design. Deshalb werden von den durchführenden Organisationen Standards für die Stichprobenziehung festgelegt. Zudem kann vorgegeben werden, wie viele Schulen in der Stichprobe enthalten sein sollten. Bei IGLU waren dies beispielsweise 150 Schulen pro Land.

Größe der  
Stichprobe

Des Weiteren wurden die Stichproben in IGLU durch ein zweistufiges Verfahren geleitet. Zunächst wurden bundeslandweise Schulen gezogen, anschließend Klassen innerhalb dieser Schulen. Die Schulen wurden mit einer Wahrscheinlichkeit proportional zur Größe der Schule gezogen. Dies ist ein Verfahren, das in den meisten Studien des Systemmonitorings angewendet wird, da die Schulanzahl von Jahr zu Jahr als konstant angesehen wird und die Schülerzahlen ebenfalls keinen großen Schwankungen unterliegen. Für alle Klassen einer Schule besteht dann die gleiche Wahrscheinlichkeit, für die Stichprobe gezogen zu werden.

Ziehung der  
Schülerinnen  
und Schüler

Durch diese Methode repräsentiert ein an der Studie teilnehmender Schüler nicht nur sich selbst, sondern auch eine bestimmte Anzahl anderer Schülerinnen und Schüler. Die Anzahl dieser Personen lässt sich dadurch berechnen, wie groß seine eigene Wahrscheinlichkeit ist, in die Stichprobe zu gelangen. Diese wiederum setzt sich aus der Wahrscheinlichkeit zusammen, dass die Schule gezogen wird und der Wahrscheinlichkeit, dass die Klasse innerhalb der Schule gezogen wird (Bos u.a., 2003). Möglich ist es aber auch, dass bestimmte Schulen oder Personen in der Stichprobe überrepräsentiert werden, ein solches Vorgehen nennt sich ‚oversampling‘. Dies war beispielsweise in TIMSS III (1995) der Fall. Da zusätzliche Untersuchungen nur für die gymnasiale Oberstufe durchgeführt werden sollten, wurden Gymnasien stärker in die Stichprobe einbezogen. Das gleiche galt für die fünf am stärksten vertretenen Ausbildungsberufe an den Berufsschulen (Baumert/Bos/Watermann, 1999). Diese Überrepräsentation muss dann in der normalen Auswertung beachtet und relativiert werden.

Wichtig ist zudem, dass die ausgewählten Schülerinnen und Schüler auch tatsächlich an der Untersuchung teilnehmen. Daher wurde beispielsweise bei PISA 2006 eine Mindestbeteiligungsquote von 80 Prozent gefordert. Trotzdem ist nicht auszuschließen, dass eher die leistungsschwachen Schülerinnen und Schüler der Untersuchung fernbleiben. Daher wurden im Rahmen von PISA 2003 und 2006 für alle gezogenen Schülerinnen und Schüler die entsprechenden Schulnoten festgehalten. Eine Analyse der Daten ergab, dass die Noten der teilnehmenden und der nicht-teilnehmenden Schülerinnen und Schüler sich nur gering unterschieden und somit ausgeschlossen werden konnte, dass eine Ergebnisverzerrung vorliegt (PISA-Konsortium, 2007).

Tatsächliche  
Teilnahme am  
Test

Soll eine nationale Ergänzungsstudie durchgeführt werden, so muss die Stichprobe erhöht werden, damit die dort vorhandene Fragestellung zufallskritisch abgesichert werden kann. Damit die Ergebnisse auf nationaler Ebene interpretiert werden können, werden die Bundesländer entsprechend der proportionalen Anteile der realen Schülerzahlen gewichtet (Bos u.a., 2003).

Ergänzungs-  
studien

### 3.3.6 Übersicht durchgeführter und anstehender Erhebungen (optional)

Die folgenden Tabellen geben einen Überblick darüber, welche Studien bereits durchgeführt wurden und welche noch durchgeführt werden. In den Tabellen ist vermerkt, welche Bezugsgruppe und welche Inhalte in den Studien getestet wurden beziehungsweise werden.

<sup>16</sup> Werden in einem Land zu viele Schülerinnen und Schüler ausgeschlossen, die durch ihre Einschränkungen tendenziell schlechtere Ergebnisse erreicht hätten, so werden die Ergebnisse überschätzt (Baumert/Lehmann, 1997).

Wann?	Wer?	Was?	Studie
1995	Sekundarstufe I und II	Mathematik und Naturwissenschaften	TIMSS
2000	Sekundarstufe I	Lesen, Mathematik, Naturwissenschaften	PISA
2001	Primarstufe	Lesen, Mathematik	IGLU
2003	Sekundarstufe I	Lesen, Mathematik, Naturwissenschaften	PISA
2006	Primarstufe	Lesen, Mathematik	IGLU
2006	Sekundarstufe I	Lesen, Mathematik, Naturwissenschaften	PISA
2007	Primarstufe	Mathematik, Naturwissenschaften	TIMSS
2009	Sekundarstufe I	Lesen, Mathematik, Naturwissenschaften	PISA

Tabelle 9: Bereits durchgeführte Erhebungen von Studien des Systemmonitorings

Wann?	Wer?	Was?	Studie
2011	Primarstufe	Deutsch	IGLU
2011	Primarstufe	Mathematik	TIMSS
2012	Sekundarstufe I	Lesen, Mathematik, Naturwissenschaften	PISA
2015	Sekundarstufe I	Lesen, Mathematik, Naturwissenschaften	PISA
2016	Primarstufe	Deutsch, Mathematik	IGLU
2018	Sekundarstufe I	Lesen, Mathematik, Naturwissenschaften	PISA

Tabelle 10: Geplante Erhebungen von Studien des Systemmonitorings

Die Tabellen zeigen den geplanten Rhythmus der jeweiligen Studien: Alle drei Jahre wird PISA durchgeführt, alle fünf Jahre IGLU, alle vier Jahre TIMSS.

#### Plöner Beschlüsse

Eine weitere wichtige Entscheidung hat die KMK 2006 in ihren „Plöner Beschlüssen“ festgehalten: Die regelmäßige Durchführung des Systemmonitorings soll mit den Bildungsstandards verbunden werden. Dazu werden die oben beschriebenen Studien eingesetzt, um zu überprüfen, wie hoch der Anteil der Schülerinnen und Schüler ist, die die Bildungsstandards erreichen (Köller, 2008). Für die Primarstufe wird in den Fächern Deutsch und Mathematik, für den Hauptschulabschluss die Fächer Deutsch, Mathematik und erste Fremdsprache und für den mittleren Schulabschluss die Fächer Deutsch, Mathematik, erste Fremdsprache, Biologie, Chemie und Physik, jeweils an die internationalen Studien angelehnt, getestet.

### 3.3.7 Weiterführende Literatur

Einen genauen Einblick in die Testkonstruktion der einzelnen Studien geben die Berichte der Studien (siehe Kapitel 3.1.3.1 bis 3.1.3.4). Nähere Informationen zum Umgang mit Multi-Matrix-Designs bietet dabei vor allem die PISA-Literatur.

### 3.3.8 Verständnis-Aufgaben und Diskussionspunkte

1. Warum sollten Aufgaben für eine Studie auf ihre Praxistauglichkeit untersucht werden?
2. Wie unterscheiden sich Aufgaben für eine Klassenarbeit von Testaufgaben?
3. Recherchieren Sie die Bedeutung des Begriffes „kumulatives Wissen“, setzen Sie ihn in den Zusammenhang mit der Kompetenzniveauentwicklung und entwickeln Sie ein Beispiel.

### 3.4 Anwendungsbereich

Kapitel 3.4 beschreibt den Anwendungsbereich von Systemmonitoring-Studien. Dazu werden die Ergebnisse der in Kapitel 3 betrachteten Studien zusammengefasst, zuvor werden die notwendigen theoretischen Grundlagen beschrieben. Zudem soll beschrieben werden, in wie weit sich durch die Ergebnisse Hinweise auf Veränderungsnotwendigkeiten im deutschen Bildungssystem erkennen lassen.

In diesem Kapitel werden folgende Fragen beantwortet:

- *Welche theoretischen Grundlagen werden für eine Interpretation der Ergebnisse benötigt?*
- *Welche Ergebnisse und Interpretationen haben die bisherigen Erhebungen von TIMSS, PISA, IGLU und DESI gezeigt?*
- *Welche Aussagekraft haben diese Ergebnisse?*
- *Wie können die Ergebnisse im Schulalltag verwendet werden?*

#### 3.4.1 Theoretische Grundlagen

Wie bereits im vorangegangenen Kapitel beschrieben, werden die Ergebnisse in den Systemmonitoringstudien skaliert, sodass alle Studien einen Mittelwert von 500 Punkten und eine Standardabweichung von 100 Punkten aufweisen.<sup>17</sup>

Skalierung

Erreicht ein Schüler einen Punktwert von 500, so entspricht dieser der mittleren Kompetenz aller Schülerinnen und Schülern in den Teilnehmerstaaten. Die Ergebnisse der Schülerinnen und Schüler sind so standardisiert, dass 68,2 Prozent und damit etwa zwei Drittel mit ihrem Punktwert zwischen den Punktwerten 400 und 600, also eine Standardabweichung über und eine Standardabweichung unter dem Mittelwert, liegen. Erhöht man dies auf jeweils zwei Standardabweichungen, so liegen zwischen einem Punktwert von 300 und 700 Punkten 95,4 Prozent der Schülerinnen und Schüler mit ihrem Punktwert in diesem Bereich. Daraus folgt, dass nur 2,3 Prozent der an der Studie teilnehmenden Schülerinnen und Schüler einen Punktwert von unter 300 oder über 700 erreichen.

Streuung

Mittelwert und Standardabweichung helfen ebenfalls dabei, internationale Vergleiche anzustellen. Liegt beispielsweise der Mittelwert eines Staats bei 600 Punkten, so kann konstatiert werden, dass die Schülerinnen und Schüler in diesem Staat Ergebnisse erreichen, die deutlich über denen des internationalen Durchschnitts liegen. Für einen Vergleich kann auch die Standardabweichung eines einzelnen Staats hilfreich sein. Weist ein Land eine Standardabweichung auf, die größer ist als 100 Punkte, so streuen die Ergebnisse der Schülerinnen und Schüler in diesem Land stärker um den Mittelwert. Wäre die Standardabweichung kleiner, so hätten die Schülerinnen und Schüler homogenere Ergebnisse erreicht.

Internationale Vergleiche

Außerdem sind die Kompetenzniveaus wichtig für die Interpretation der Ergebnisse. Sie können dafür genutzt werden, um zu überprüfen, in wie weit sich die Verteilung auf die einzelnen Stufen in den Staaten unterscheidet. Zudem kann so der Anteil der Schülerinnen und Schüler ausgemacht werden, die schlechte Chancen für ihren weiteren Lebensweg haben, da sie sich auf dem untersten Kompetenzniveau oder sogar noch darunter befinden. Dies gilt auch umgekehrt für die Schülerinnen und Schüler, die besonders gut abschneiden und besonders gute Chancen haben (PISA-Konsortium Deutschland, 2007). Eine Betrachtung der Kompetenzniveaus bietet somit einen höheren Informationsgehalt als reine Vergleiche der Mittelwerte (Deutsches PISA-Konsortium, 2001).

Kompetenz-niveaus

#### 3.4.2 Wesentliche Ergebnisse (optional)

In diesem Kapitel sollen die Ergebnisse der Studien TIMSS, PISA, IGLU und DESI vorgestellt werden. Eine solche Betrachtung bisheriger Ergebnisse verdeutlicht, wie zukünftige Ergebnisse interpretiert werden können und wie groß das Spektrum möglicher Interpretationen ist. Dabei sollen nicht nur reine Punktwerte sondern auch mögliche Erklärungen für diese Werte beschrieben werden. Das Hauptaugenmerk soll auf den internationalen Vergleichen liegen und dabei auf den Ergebnissen der in Deutschland getesteten Schülerinnen und Schüler. Da aber auch die Ergebnisse der Ergänzungsstudien wichtige Aussagen über Veränderungsnotwendigkeiten innerhalb Deutschlands liefern, sollen diese jeweils kurz betrachtet werden. Interessante Aspekte aus der Erhebung von Hintergrundinformationen (wie etwa zum sozialen Hintergrund der Schülerinnen und Schüler) sollen ebenfalls betrachtet werden. Da der Studienbrief-Teil nicht allumfassend über die Ergebnisse berichten kann und dies auch nicht seinem Ziel entspricht, sollen nur allgemeine und besonders auffällige Aspekte beschrieben werden. Dabei soll deutlich werden, dass nicht die reine Rangfolge, wie in der Presse häufig verwendet, wichtig ist, sondern vielmehr die jeweiligen erreichten Werte wie Mittelwert, Standardabweichung und die Verteilung auf die Kompe-

<sup>17</sup> Bei einer wiederholten Durchführung der Studie mit denselben Test-Items wird dann der tatsächlich errechnete Wert als Standard verwendet.



tenzniveaus. Weitere Informationen zu den einzelnen Studien können der jeweils angegebenen Literatur entnommen werden.

### 3.4.2.1 TIMSS (optional)

Generell kann für die Leistungen der deutschen Schülerinnen und Schüler in TIMSS 1995 festgehalten werden, dass sie sowohl in Mathematik als auch in den Naturwissenschaften „in einem breiten internationalen Mittelfeld“ liegen. Dies lässt sich sowohl für die Sekundarstufe I (TIMSS II) als auch für die Sekundarstufe II (TIMSS III) erkennen:

- |           |  |
|-----------|--|
| TIMSS II  | <ul style="list-style-type: none"> <li>• TIMSS II: Deutsche Schülerinnen und Schüler erreichen einen Mittelwert von 509 Punkten, wobei der internationale Durchschnitt bei 513 Punkten liegt, das bedeutet, dass beispielsweise mathematische Routineverfahren aus der sechsten bis achten Klasse „einigermaßen sicher“ durchgeführt werden können.</li> <li>• TIMSS II: Leistungen in den Naturwissenschaften besser als in Mathematik, allerdings nicht so große Unterschiede wie in anderen Staaten</li> <li>• Besorgniserregend: Etwa 20 Prozent der Schülerinnen und Schüler in der achten Klasse sind im Bereich der Naturwissenschaften noch auf dem Niveau von Grundschulkenntnissen.</li> </ul> |
| TIMSS III | <ul style="list-style-type: none"> <li>• TIMSS III: Defizite auch in der gymnasialen Oberstufe</li> <li>• Für alle Schulformen: Auch bei den Schülerinnen und Schülern der Sekundarstufe II ist im naturwissenschaftlichen Bereich das unterste Kompetenzniveau überrepräsentiert.</li> <li>• Defizite besonders bei der Verknüpfung der fachlichen Kenntnisse mit der Anwendung bei alltagsnahen Problemen. Dies lässt auf eine im Unterricht fehlende Anbindung an die Alltagswelt der Schüler schließen.</li> </ul>   |

Zur Veranschaulichung der Ergebnisse wird eine Tabelle dargestellt, wie sie in den Berichten der Studien meist vorhanden ist. Dabei wird für die vier Kompetenzniveaus (hier in ihrer ausformulierten Form von oben nach unten, für den Bereich Mathematik in TIMSS III) angegeben, wie viel Prozent der Schülerinnen und Schüler diese erreichen.

	Deutschland	Frankreich	Niederlande	Norwegen	Schweiz
Alltagsbezogene Schlussfolgerungen	15,4	0,0	3,7	0,6	0,8
Anwendung von einfachen Routinen	36,6	34,3	21,5	35,0	29,3
Bildung von Modellen und Verknüpfung von Operationen	34,1	47,6	41,3	40,2	43,0
Mathematisches Argumentieren	13,9	18,1	33,4	24,2	26,9

Tabelle 11: Verteilung der Schülerinnen und Schüler aus Deutschland, Frankreich, Niederlande, Norwegen und der Schweiz auf die vier Kompetenzniveaus des Bereichs Mathematik

Hier wird deutlich, dass Deutschland im untersten Niveau überrepräsentiert (15,4 Prozent) und im obersten Niveau unterrepräsentiert ist (13,9 Prozent).

Im internationalen Vergleich von TIMSS II **wurde** die folgenden Aspekte deutlich:

- |                                |   |
|--------------------------------|---|
| Schüler älter                  | <ul style="list-style-type: none"> <li>• Die deutschen Schülerinnen und Schüler erreichen das Niveau der internationalen Mittelgruppe (Deutschland und 11 weitere Staaten mit ähnlichen Ergebnissen) erst, wenn sie sechs bis zwölf Monate älter sind als die Schülerinnen und Schüler aus den anderen Staaten dieser Gruppe (durch spätere Einschulungszeitpunkte, Zurückstellungen und eine hohe Quote von Wiederholern), diese Entwicklung zeigte sich auch in TIMSS III.</li> </ul> |
| Vergleichbare Länder           | <ul style="list-style-type: none"> <li>• Am besten kann Deutschland in TIMSS II mit den angelsächsischen Staaten verglichen werden, da die Schülerinnen und Schüler der nord-, ost- und westeuropäischen sowie asiatischen Staaten einen Leistungsvorsprung von mehr als einem Jahr aufweisen können.</li> </ul>  |
| Leistungsgruppe Mathematik     | <ul style="list-style-type: none"> <li>• Im Bereich der Mathematik bilden diese Staaten eine Leistungsgruppe, die Punktwerte erreichen konnte, die etwa eine halbe Standardabweichung über den deutschen Ergebnissen liegt, dies entspricht in etwa dem Leistungsfortschritt aus einem Schuljahr.</li> </ul>  |
| Entwicklung über ein Schuljahr | <ul style="list-style-type: none"> <li>• Von den besten zu den schlechtesten Schülerinnen und Schülern besteht ein großer Unterschied von mehreren Schuljahren, sodass dieser Vorsprung praktisch nicht mehr eingeholt werden kann. Dies wird vor al-</li> </ul>  |

lem dadurch verstärkt, dass sich die Unterschiede zwischen den guten asiatischen Staaten vom Beginn der Sekundarstufe I zum Ende der Sekundarstufe I beispielsweise im Bereich Mathematik von knapp einer Standardabweichung auf über eine Standardabweichung nochmals vergrößern (auch durch TIMSS III wird erkennbar, dass der Abstand größer wird)

In Bezug auf die Streuung der Punktwerte lassen sich die folgenden Aspekte zusammenfassen:

Streuung

- Streuung der Punktwerte bei deutschen Schülerinnen und Schülern besonders groß, große Heterogenität in beiden getesteten Bereichen, mit 90 Punkten allerdings im Bereich der Mathematik in der Nähe des internationalen Mittels
- Naturwissenschaften: Standardabweichung bei 101 Punkten, damit einerseits größer als in Mathematik, andererseits deutlich über dem Wert der anderen teilnehmenden Staaten
- Streuung von Kompetenzen der fünften bis zur zehnten Klasse, sogar innerhalb einer einzelnen Schulform Leistungsunterschiede von über zwei Schuljahren
- Streuung in den Naturwissenschaften geringer als in Mathematik
- Homogene Leistungsbilder nur in Staaten, die insgesamt schwächere Leistungen erbringen (Beispiel: Portugal, Iran)

Zwar gibt es auch einige deutsche Schülerinnen und Schüler, die Spitzenleistungen im Bereich Mathematik erreichen, dies sind allerdings nur sehr wenige. International erreichen in TIMSS II 10 Prozent der Schülerinnen und Schüler einen Punktwert von über 656 Punkten. Somit müssten auch in Deutschland 10 Prozent diesen Wert erreichen oder übertreffen, es sind aber nur 6 Prozent. TIMSS III stellte fest, dass die leistungsstärksten Schülerinnen und Schüler aus Deutschland nicht mit den starken Schülerinnen und Schülern aus anderen europäischen Staaten mithalten können.

Spitzenleistungen

Die Ergänzungsstudie zu TIMSS 1995 zeigte, dass Schülerinnen und Schüler aus den neuen Bundesländern bessere Ergebnisse erreichen. Diese sind wahrscheinlich darauf zurückzuführen, dass dort Mädchen besser an gymnasialer Bildung beteiligt und leistungsschwächere Schülerinnen und Schüler besser gefördert wurden.

Ergänzungsstudie

Weitere Erkenntnisse aus der Ergänzungsstudie:

- Innerhalb der verschiedenen Schulformen konnte festgestellt werden, dass sich die Fähigkeiten überlappen, 30 Prozent der Realschüler und 25 Prozent der Gesamtschüler liegen mit ihren Fähigkeiten oberhalb des durchschnittlichen Niveaus der Gymnasiasten.
- Vor allem in der Realschule ließ sich eine große Variabilität feststellen. Generell konnte in beiden Fachbereichen erkannt werden, dass die unterschiedlichen Schulformen auch unterschiedliche Ergebnisse erreichen.
- Leistungsfortschritte, die Schülerinnen und Schüler in Deutschland von der siebten zur achten Klasse erreichen, sind im internationalen Vergleich nur gering.

Schulformen

In allen Schulformen und in beiden untersuchten Fachrichtungen erreichen Mädchen schlechtere Leistungen als Jungen, vor allem in Physik sind die Unterschiede besonders groß (Unterschied in den neuen Bundesländern aber deutlich geringer als in den alten Bundesländern).

Mädchen / Jungen

Durch die Fragebögen wurde in TIMSS II deutlich, dass die Schule bei der Identitätsbildung der Schülerinnen und Schüler behilflich sein und den Prozess des Erwachsenwerdens stabilisieren kann. Die ‚Schulunlust‘ bleibt innerhalb der Untersuchungen von TIMSS unverändert und unterscheidet sich auch nicht in Geschlecht oder Schulform. Das Interesse an den Gegenständen des Unterrichts nimmt innerhalb des Untersuchungszeitraumes ab. Wichtige Erkenntnis ist, dass das Selbstwertgefühl der Schülerinnen und Schüler im achten Schuljahr sinkt, das Vertrauen in die schulische Leistungsfähigkeit allerdings stabil bleibt. Dabei gibt es deutliche Unterschiede zwischen Mädchen und Jungen. Selbstzweifel und Leistungsängste sind bei Mädchen verstärkt dokumentiert. Dies ist eine Erkenntnis, die international gültig ist. Dabei hängen Selbstzweifel und Leistungsängste nicht mit schlechteren Leistungen in den Fächern zusammen.

Fragebögen

TIMSS 2007

2007 wurden in Deutschland im Rahmen von TIMSS nur Grundschülerinnen und Grundschüler am Ende der vierten Klasse. Die Ergebnisse im Vergleich zum internationalen Mittelwert zeigt folgende Tabelle 12:

	Internationaler Durchschnitt	Deutscher Mittelwert
<b>Mathematik</b>	473 Punkte	525 Punkte
<b>Naturwissenschaften</b>	476 Punkte	528 Punkte

Tabelle 12: Internationaler und deutscher Durchschnitt bei TIMSS 2007 im Vergleich

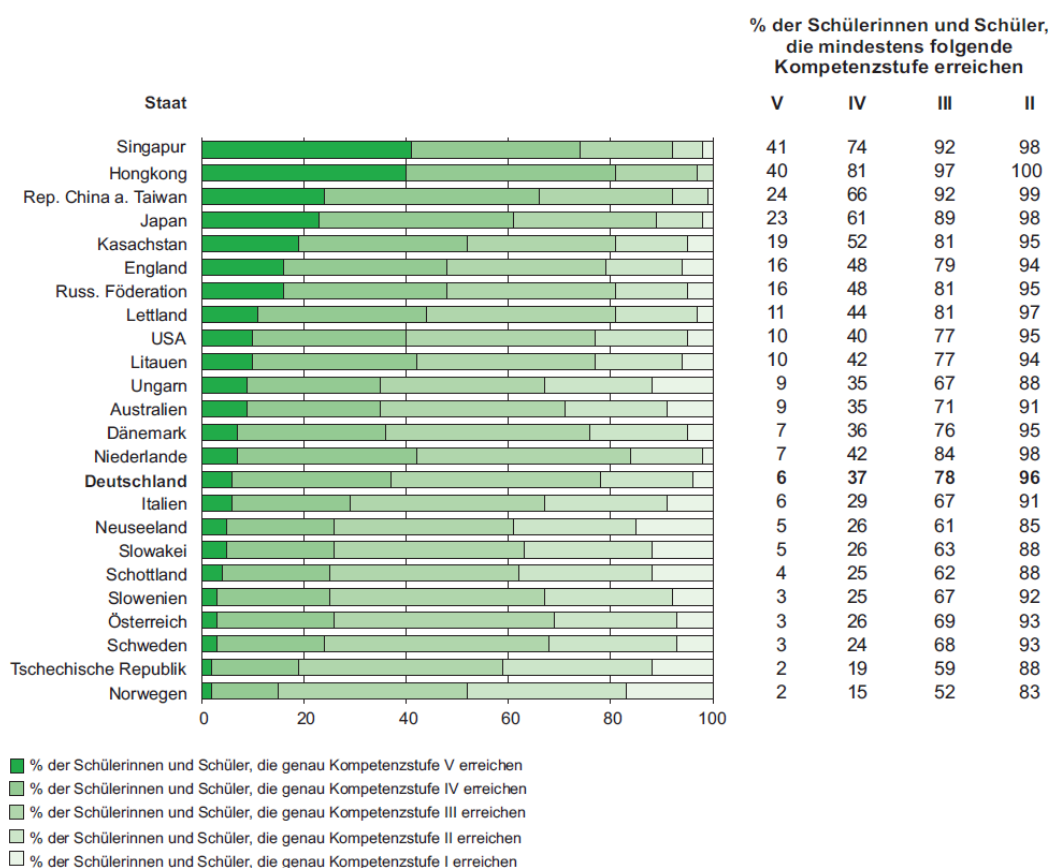
Es besteht allerdings kaum Abstand zu den anderen europäischen Staaten, dagegen gibt es große Unterschiede zu den besser abscheidenden Staaten wie beispielsweise Japan. Leider konnten keine Langzeitentwicklungen aufgezeigt werden, da Deutschland sich zum ersten Mal beteiligte.

Streuung

Zudem konnte für die deutschen Schülerinnen und Schüler eine geringere Leistungsstreuung festgestellt werden, da die Standardabweichung vom deutschen Mittelwert in Mathematik bei 68 Punkten lag. Ein solch homogenes Bild lässt sich ebenfalls für die Naturwissenschaftsleistungen feststellen.

Kompetenz-  
niveaus  
Mathematik

Leider zeigten auch die Ergebnisse aus 2007 eine hohe Anzahl von Schülerinnen und Schülern mit geringen mathematischen Kompetenzen. So befindet sich etwa ein Fünftel auf den Kompetenzniveaus I und II. Zudem existiert ein Anteil von vier Prozent der Kinder, die nur mangelhafte Leistungen auf dem untersten Kompetenzniveau erreichen. Auf dem höchsten Kompetenzniveau gibt es sechs Prozent der Schülerinnen und Schüler, die mit ihren mathematischen Fähigkeiten komplexe Probleme lösen können. Diese Verteilung wird von Abbildung 5 nochmals verdeutlicht:



IEA: Trends in International Mathematics and Science Study

© TIMSS 2007

Abbildung 5: Verteilung der teilnehmenden Staaten auf die Kompetenzniveaus des Bereichs Mathematik

Bei den Naturwissenschaften befindet sich etwa ein Viertel der Schülerinnen und Schüler auf den beiden untersten Kompetenzniveaus, sechs Prozent erreichen nur rudimentäre Kenntnisse und liegen unter dem untersten Niveau. Das höchste Niveau wird von zehn Prozent im naturwissenschaftlichen Bereich erreicht. Sowohl die Verkleinerung des Anteils der Schülerinnen und Schüler auf der untersten Stufe als auch die Vergrößerung des Anteils der Schülerinnen und Schüler auf der obersten Stufe wird von vielen anderen Staaten, beispielsweise Japan, wesentlich besser gemeistert.

Naturwissenschaften

Kompetenzniveaus

Bei der Frage nach einem geschlechtsspezifischen Unterschied konnte Folgendes festgestellt werden:

Mädchen / Jungen

- Im Bereich der Mathematik lassen sich international keine geschlechtsspezifischen Unterschiede feststellen, Jungen und Mädchen erreichen beide einen Mittelwert von 473.
- In einem Drittel der teilnehmenden Staaten haben die Jungen einen deutlichen Vorsprung in Mathematik, Deutschland gehört dazu (Schüler: 531 Punkte, Schülerinnen: 519 Punkte).
- Naturwissenschaften: International zeigen sich ebenfalls keinerlei geschlechtsspezifische Unterschiede.
- Deutschland ist das einzige Land unter den EU- beziehungsweise OECD-Staaten, in dem es einen großen Unterschied in den Naturwissenschaften zwischen Mädchen und Jungen gibt
- Die deutschen Grundschülerinnen und Grundschüler haben eine positive Einstellung der Mathematik gegenüber und liegen damit im internationalen Trend. Dabei ist zu erkennen, dass die Einstellung dem Fach gegenüber unabhängig von der Leistung ist. Bei 81 Prozent aller Grundschülerinnen und Grundschüler ist die Einstellung gegenüber den Naturwissenschaften positiv.

Einstellungen

(Baumert/Lehmann, 1997; Baumert/Bos/Watermann, 1999; Köller/Baumert/Bos, 2001; Bos/Bonsen/Baumert/Prenzel/Selter/Walther, 2008)

### 3.4.2.2 PISA (optional)

Im Jahr 2000 bildet die Lesekompetenz den Schwerpunkt der Untersuchung. Die Ergebnisse sollen stichpunkthaft zusammengefasst werden:

PISA 2000

- Deutschland liegt im Schwerpunktbereich Lesen mit einem Mittelwert von 484 Punkten 16 Punkte unter dem internationalen Mittelwert<sup>18</sup>, damit gehört Deutschland zu den 14 Staaten, die sich mit ihrem Mittelwert signifikant unter dem internationalen Mittelwert befinden.
- Im internationalen Vergleich ist der Anteil der Schülerinnen und Schüler, die unter dem untersten Kompetenzniveau liegen, besonders groß (international: sechs Prozent, Deutschland: zehn Prozent), viele andere Staaten (z.B. Australien und Finnland) weisen hier einen Anteil von unter fünf Prozent auf.
- In Deutschland befinden sich nur 12,7 Prozent der Schülerinnen und Schüler auf Kompetenzniveau I, fast 23 Prozent (international: 18 Prozent) liegen unter Kompetenzniveau II und können damit elementare Aufgaben nicht mit entsprechender Sicherheit lösen.
- In Deutschland liegen neun Prozent der Schülerinnen und Schüler auf dem obersten Kompetenzniveau (V) (internationaler Durchschnitt: 9,5 Prozent), damit liegt Deutschland im guten Mittelfeld. Allerdings erreichen die Spitzen wie Neuseeland und Finnland mit über 18 Prozent ein deutlich besseres Ergebnis.

Lesen

Kompetenzniveaus

Im Vergleich zu anderen Staaten ist die Spannweite der Ergebnisse der deutschen Schülerinnen und Schüler im Bereich Lesen besonders groß. Der Leistungsabstand zwischen den fünf Prozent der besten und den fünf Prozent der schlechtesten Schülerinnen und Schüler ist im internationalen Vergleich der Größte. Hier ergibt sich ein Unterschied von 366 Punkten, der damit um 38 Punkte größer ist, als der internationale Durchschnitt der Streuung. Entsprechend diesem Befund weist Deutschland mit 111 Punkten auch die größte Standardabweichung auf.

Streuung

Der Mittelwert von Deutschland im Bereich Mathematik liegt mit 490 Punkten zehn Punkte unter dem internationalen Mittelwert und innerhalb der unteren Hälfte des Mittelfeldes, gemeinsam mit den Vereinigten Staaten, Spanien und den osteuropäischen Staaten.

Mathematik

Tabelle 13 zeigt die Verteilung der deutschen Schülerinnen und Schüler auf die Kompetenzniveaus:

Kompetenzniveaus

Unter I	I	II	III	IV	V
7 %	17 %	32 %	31 %	12 %	1,3 %

Tabelle 13: Verteilung der deutschen Schülerinnen und Schüler auf die Kompetenzniveaus des Bereichs Mathematik bei PISA 2000

<sup>18</sup> Unter dem internationalen Mittelwert wird im Folgenden der in den Berichten zu den Studien angegebene OECD-Mittelwert verstanden.

Der Anteil derjenigen Schülerinnen und Schüler, die unter Kompetenzniveau I liegen, ist in Deutschland am größten, doppelt so groß wie in Großbritannien und dreimal so groß wie in Japan. Im internationalen Vergleich setzt sich besonders Japan ab, da hier der Schwerpunkt auf den Kompetenzniveaus III bis V liegt. Hier ist der Anteil derer, die die höchste Stufe im Bereich Mathematik erreichen, sogar viermal so groß wie in Deutschland. Allerdings schneiden im oberen Bereich Staaten wie Norwegen noch schlechter ab als Deutschland.

Die Standardabweichung liegt in Mathematik bei 103 Punkten und ist damit kleiner als im Bereich Lesen, allerdings noch immer größer als die internationale Standardabweichung von 100 Punkten. Zudem ist Deutschland auch hier wieder unter den Staaten mit der größten Streuung der Ergebnisse. Vergleicht man die Punktwerte der besten und schlechtesten fünf Prozent der Schüler, so ergibt sich ein Punktunterschied von 338 Punkten, also mehr als drei Standardabweichungen. Ähnliche Werte ergeben sich nur in wenigen weiteren Staaten.

Bei den Naturwissenschaften zeigen sich die folgenden Ergebnisse:

- Der Mittelwert für die deutschen Schülerinnen und Schüler liegt bei 487 Punkten und somit 13 Punkte unter dem international durchschnittlichen Mittelwert von 500 Punkten.
- Die besten Ergebnisse in diesem Bereich erreicht Korea mit 552 Punkten und liegt damit sehr weit vor Deutschland.
- Da Brasilien als schlechtestes Land nur 375 Punkte erreicht, wird deutlich, wie weit die Ergebnisse international auseinander liegen.
- Die Verteilung auf die Kompetenzniveaus liegt in Deutschland anders vor, als dies international der Fall ist: Mehr Schülerinnen und Schüler liegen auf den unteren Kompetenzniveaus (52,6 Prozent auf Stufe I und II im Vergleich zu 50,1 Prozent international) und weniger auf den hohen (27,3 Prozent auf Stufe IV und V im Vergleich zu 29,8 Prozent international).
- Die Standardabweichung liegt für die deutschen Schülerinnen und Schüler bei 102 Punkten und somit nicht weit entfernt vom internationalen Wert von 100 Punkten, gerade bei den Staaten mit sehr guten Ergebnissen ist eine wesentlich geringere Streuung der Ergebnisse vorhanden.

Die Ergebnisse der Ergänzungsstudie zeigen:

- Die meisten Bundesländer liegen unter dem internationalen Durchschnitt.
- Die Länder unterscheiden sich untereinander immens, es gibt Länder, die im oberen Drittel und solche, die sich am Ende der internationalen Rangreihe befinden.
- Für den Bereich Lesen zeigt sich in allen Bundesländern, dass die schwächsten Schülerinnen und Schüler besonders niedrige Leistungen im internationalen Vergleich erbringen, zudem haben alle Länder außer Bayern einen erhöhten Anteil von Risikoschülern, die nicht über die untersten Kompetenzniveaus hinaus kommen.
- Im Bereich Mathematik ergibt sich über die Bundesländer hinweg eine besonders große Streuung von 64 Punkten (Unterschied Bremen zu Bayern), zudem erreicht kein Bundesland ein Ergebnis, das mit der internationalen Spitzengruppe mithalten kann.
- Auch in den Naturwissenschaften ist die Streuung von 461 Punkten (Bremen) zu 508 Punkten (Bayern) sehr groß, die Leistungen innerhalb der Bundesländer unterscheiden sich also stark.
- Auch die Leistungen der besten deutschen Bundesländer sind im internationalen Vergleich nur im Mittelfeld vorzufinden.

Für den Bereich Lesen ergeben sich in Deutschland Hinweise auf einen Zusammenhang zwischen den Ergebnissen und der Zeit, die die Schülerinnen und Schüler freiwillig mit Lesen verbringen. In Deutschland liegt der Anteil der Schülerinnen und Schüler, die überhaupt nicht zum Vergnügen lesen, bei 42 Prozent.

Streuung

Naturwissenschaften

Kompetenzniveaus

Streuung

Ergänzungsstudie

Lesen

Mathematik

Naturwissenschaften

Interessante Aspekte

Zum Thema ‚Geschlechterunterschied‘ lassen sich folgende Ergebnisse aufzeigen:

- Der Unterschied ist international beim Lesen am größten, die mittlere Geschlechterdifferenz beträgt hier 32 Punkte, in Deutschland liegt sie mit 35 Punkten nah an diesem gemittelten Wert.
- Mädchen haben gegenüber Jungen einen Vorsprung von knapp einem halben Kompetenzniveau, allerdings gibt es Staaten, wie beispielsweise Finnland, die generell sehr gut abgeschnitten haben, wo die Geschlechterdifferenz jedoch mit 51 Punkten sehr hoch ist.
- In Mathematik zeigte sich ein Vorsprung bei den Jungen, allerdings ist dieser nicht so groß wie der der Mädchen beim Lesen (11 Punkte im internationalen Durchschnitt, 15 Punkte in Deutschland).
- Im Bereich Naturwissenschaften gibt es nur wenige Staaten, in denen sich Unterschiede zwischen Mädchen und Jungen zeigen, Deutschland gehört nicht dazu.

Es wurde festgestellt, dass in keinem anderen Industrieland außer Deutschland der Zusammenhang zwischen sozialer Herkunft und fachlichen Fähigkeiten so groß ist wie in Deutschland. Vor allem Schülerinnen und Schüler mit Migrationshintergrund, in deren Familien kaum Deutsch gesprochen wird, bleiben auf den Kompetenzniveaus deutlich hinter den Anderen. Zudem konnte auch hier, wie bereits bei TIMSS, festgestellt werden, dass die Schülerinnen und Schüler, die mit 15 Jahren getestet wurden, ein bis zwei Jahre hinter Schülerinnen und Schüler anderer Staaten zurückliegen.

Zusammen-  
hang mit  
sozialer  
Herkunft

Deutschland liegt beim PISA 2003-Schwerpunktthema Mathematik mit einem Mittelwert von 503 Punkten sehr nah am internationalen Mittelwert von 500 Punkten. Das beste Ergebnis im Bereich Mathematik erreicht Finnland mit 544 Punkten, ein großer Unterschied zum deutschen Ergebnis wird dabei deutlich.

PISA 2003  
Mathematik

Die Verteilung auf die Kompetenzniveaus zeigt Tabelle 14:

Unter I	I	II	III	IV	V	VI
9,2 %	12,4 %	19 %	22,6 %	20,6 %	12,2 %	4,1 %

Tabelle 14: Verteilung der deutschen Schülerinnen und Schüler auf die Kompetenzniveaus des Bereichs Mathematik bei PISA 2003

Kompetenz-  
niveaus

Dabei ist zu erkennen, dass der Anteil der Schülerinnen und Schüler, die als Risikogruppe gelten, sehr groß ist, er liegt bei 21,6 Prozent. Dieser Wert ist der zweithöchste im internationalen Vergleich, der Abstand zu den anderen west- und nordeuropäischen Staaten ist sehr groß. Bei der Verteilung auf das höchste Niveau (VI) liegt Deutschland mit einem Wert von 4,1 Prozent beim internationalen Durchschnitt (4,0 Prozent). Andere Staaten (Beispiel: Belgien) erreichen hier allerdings einen Wert von 9 Prozent.

Die Streuung fällt im internationalen Vergleich sehr unterschiedlich aus. Deutschland hat mit einer Standardabweichung von 102 Punkten den dritthöchsten Wert, nach Belgien und der Türkei.

Streuung

Im Bereich Mathematik lässt sich von 2000 zu 2003 eine Verbesserung in den mathematischen Kompetenzen der Schülerinnen und Schüler in Deutschland erkennen.

Vergleich  
2000/2003

Im Bereich Lesen liegt Deutschland mit einem Mittelwert von 491 Punkten nicht weit vom internationalen Mittelwert (494 Punkte) entfernt. Auch in diesem Bereich ist der Abstand zu den besten Staaten groß, hier liegt wiederum Finnland mit 543 Punkten an der Spitze.

Lesen

Bemerkenswert ist dabei, dass gerade die Staaten mit den höchsten Ergebnissen gleichzeitig auch die niedrigsten Standardabweichungen haben. Staaten wie Finnland und Korea schaffen es, eine hohe Leistung mit geringer Streuung zu erreichen. In Deutschland liegt die Streuung mit 109 Punkten dagegen über dem internationalen Mittel von 100 Punkten.

Streuung

Das höchste Kompetenzniveau wird im internationalen Durchschnitt von 8,3 Prozent der Schülerinnen und Schüler erreicht. Deutschland erreicht 9,6 Prozent und liegt somit über diesem Wert, allerdings erreichen Staaten wie Neuseeland, Finnland und Australien wesentlich höhere Werte. Leider liegt Deutschland auch bei der Risikogruppe mit einem Wert von 22,3 Prozent über dem internationalen Wert (19,1 Prozent). Dieser große Anteil von Schülerinnen und Schülern, die keine einfachen Leseaufgaben bewältigen können, lag bereits im Jahr 2000 auf diesem Niveau.

Kompetenz-  
niveaus

Dabei ist festzustellen, dass in allen teilnehmenden Staaten Mädchen bessere Leistungen erbringen als Jungen. Der Unterschied liegt im internationalen Durchschnitt bei 34 Punkten, auch in Deutschland ist er mit 42 Punkten sehr groß.

Mädchen /  
Jungen



Naturwissenschaften	In den Naturwissenschaften erreichen die deutschen Schülerinnen und Schüler einen Mittelwert von 502 Punkten und befinden sich damit sehr nah am internationalen Mittelwert (500 Punkte). Damit ist eine Verbesserung von 15 Punkten gegenüber dem Ergebnis von PISA 2000 zu erkennen, und das Ergebnis liegt damit nicht mehr unter dem Durchschnitt.
Streuung	Die Streuung in den Naturwissenschaften ist in Deutschland sehr groß, sie liegt bei 111 Punkten und bildet damit, gemeinsam mit Frankreich, den höchsten Wert aller teilnehmenden Staaten. Sie hat sich gegenüber PISA 2000 noch vergrößert.
Ergänzungsstudie	<p>Durch die Ergänzungsstudie wurde festgestellt:</p> <ul style="list-style-type: none"> <li>• Mehrere Bundesländer haben sich im Vergleich zum internationalen Durchschnitt verbessert.</li> <li>• Drei Länder liegen sogar über dem internationalen Durchschnitt.</li> <li>• Obwohl eine Verbesserung deutlich erkennbar ist, zeigen in der Hälfte der teilnehmenden Bundesländer die Schülerinnen und Schüler noch immer Schwächen beim Lesen.</li> <li>• Trotzdem zeigt sich, dass in vielen Bundesländern und in vielen Kompetenzbereichen eine wesentliche Verbesserung der Kompetenzniveaus stattgefunden hat, teilweise sogar mit sehr deutlichen Steigerungen.</li> <li>• Steigerungen lassen sich aber nicht nur bei den Bundesländern feststellen, die schlecht abgeschnitten haben, sondern auch bei denen, die bereits gute Ergebnisse aufzuweisen hatten.</li> </ul>
PISA 2006 Naturwissenschaften	<p>Die Ergebnisse der Naturwissenschaften als Schwerpunktbereich in der Untersuchung 2006 sollen als Erstes betrachtet werden:</p> <ul style="list-style-type: none"> <li>• International liegen die Mittelwerte in einem Bereich von 410 Punkten (Mexiko) bis 563 Punkten (Finnland).</li> <li>• Deutsche Schülerinnen und Schüler liegen mit einem Punktwert von 516 Punkten signifikant über dem internationalen Mittelwert, somit ist ein deutlicher Zuwachs in diesem Bereich seit der Untersuchung im Jahr 2003 erkennbar.</li> <li>• Trotz dieser Verbesserung gibt es noch einige Staaten, die weitaus bessere Ergebnisse vorweisen können, sodass eine Steigerung der deutschen Ergebnisse möglich ist. Gerade das Ergebnis aus Finnland zeigt, dass die Schülerinnen und Schüler dort den deutschen Schülerinnen und Schülern 1,5 bis 2 Jahre voraus sind.</li> <li>• Der Anteil der Schülerinnen und Schüler, die in den Naturwissenschaften nicht über das unterste Kompetenzniveau hinaus kommen, liegt in Deutschland bei 15,4 Prozent und damit unter dem internationalen Durchschnitt, der bei 19,2 Prozent liegt.</li> <li>• Spitzenleistungen: In Deutschland erreichen 11,8 Prozent der Schülerinnen und Schüler einen Punktwert, der sie auf die beiden obersten Kompetenzniveaus (V und VI) einteilt, wobei der internationale Durchschnittswert nur bei 9 Prozent liegt.</li> <li>• Für die Naturwissenschaften zeigt sich, dass durchgeführte Veränderungen innerhalb der Schulen langsam greifen, allerdings immer noch ‚Luft nach oben‘ ist.</li> <li>• Es zeigt sich, dass über die Jahre die Ergebnisse der schlechten Schülerinnen und Schüler in Deutschland in den Naturwissenschaften durchaus besser werden, die schlechtesten fünf Prozent erreichten im Jahr 2000 einen Wert von 314 Punkten, dieser Wert lag 2006 dagegen schon bei 345 Punkten.</li> </ul>
Streuung	Die Streuung in Deutschland ist in den Naturwissenschaften mit einer Standardabweichung von 100 Punkten noch immer groß, allerdings ist auch hier eine Verbesserung erkennbar. Es gibt aber noch immer Staaten, die wesentlich homogenere Ergebnisse erreichen, wie beispielsweise Finnland.
Lesekompetenz	<p>Bei der Betrachtung der Lesekompetenz muss bei PISA 2006 darauf geachtet werden, dass der internationale Mittelwert nicht bei 500 Punkten sondern bei 492 liegt. Die Ergebnisse waren hier:</p> <ul style="list-style-type: none"> <li>• Deutschland liegt mit einem Mittelwert in der Lesekompetenz von 495 Punkten sehr knapp über dem internationalen Durchschnitt.</li> <li>• Finnland und Korea erreichen Werte von 547 beziehungsweise 556 Punkten und schneiden deutlich besser ab als Deutschland, auch andere Staaten erreichen wesentlich bessere Ergebnisse als Deutschland.</li> <li>• International sind es bei der Lesekompetenz 20,1 Prozent der Schülerinnen und Schüler, die sich unter oder auf dem untersten Kompetenzniveau mit ihrem Punktwert befinden, somit liegt Deutschland mit 20 Prozent sehr genau am internationalen Mittel.</li> </ul>
Kompetenzniveaus	

- Spitzenbereich: 9,9 Prozent der deutschen Schülerinnen und Schüler erreichen das höchste Kompetenzniveau, im Vergleich zu 8,6 Prozent im internationalen Vergleich.

Die Standardabweichung reicht beim Lesen international von 81 Punkten (Finnland) bis zu 112 Punkten (Deutschland). Hier wird deutlich, dass die Streuung in Deutschland im internationalen Vergleich besonders groß ist. Spitzenleistungen können in diesem Bereich aber auch mit einer homogenen Ergebniszusammensetzung erreicht werden, wie das Beispiel Finnland zeigt.

Über die Jahre hinweg lässt sich eine Steigerung der Leistungen der deutschen Schülerinnen und Schüler beim Lesen erkennen, die allerdings nicht statistisch abgesichert werden kann. Außerdem hat sich der Abstand zu den Spitzenländern nicht verändert, er liegt noch immer bei 50 Punkten. Allerdings zeichnet sich in Deutschland eine verbesserte Verteilung auf die Kompetenzniveaus ab. So konnten die prozentualen Anteile auf den niedrigen Niveaus verringert und auf den oberen Niveaus vergrößert werden.

Entwicklung  
über Jahre

In der Mathematik liegt der internationale Mittelwert bei 498 Punkten, die Ergebnisse stellen sich wie folgt dar:

Mathematik

- 13 Staaten können den Mittelwert überschreiten, teilweise sogar deutlich, wie beispielsweise Finnland mit einem Mittelwert von 548 Punkten, zudem erreichen die asiatischen Staaten ebenfalls sehr hohe Werte. Diese Staaten erzielen Ergebnisse, die ein Kompetenzniveau über dem internationalen Mittelwert liegen
- Deutschland erreicht einen Mittelwert von 504 Punkten und liegt damit über dem internationalen Mittelwert, aber trotzdem sehr weit von den Spitzenländern entfernt
- Bei der Verteilung auf die Kompetenzniveaus schneidet Deutschland im internationalen Vergleich besser als der Durchschnitt, auf oder unter dem untersten Kompetenzniveau liegen international 21 Prozent der Schülerinnen und Schüler, in Deutschland sind es 19,9 Prozent
- es gibt Staaten, in denen dieser Anteil wesentlich höher ist, 56,5 Prozent der mexikanischen Schülerinnen und Schüler erreichen im Bereich Mathematik nur das unterste Kompetenzniveau beziehungsweise bleiben sogar noch darunter
- Im höchsten Kompetenzniveau (VI) schneidet Deutschland besser ab als der internationale Durchschnitt, dieser liegt bei 3,3 Prozent, in Deutschland können aber mehr Schülerinnen und Schüler (4,5 Prozent) dieses Niveau erreichen, den höchsten Wert mit 9,1 Prozent erreicht Finnland

Kompetenz-  
niveaus

Auch die Streuung unterscheidet sich teilweise deutlich von Land zu Land, teilweise sogar sehr deutlich. Die größte Standardabweichung und damit die größte Streuung ergibt sich in Belgien mit 106 Punkten. Deutschland liegt mit 99 Punkten noch signifikant über dem internationalen Mittel von 92 Punkten.

Streuung

Bei der Betrachtung der Entwicklung über die Jahre ergeben sich für den Mathematik-Bereich sowohl im internationalen Durchschnitt als auch für Deutschland kaum Veränderungen von 2003 zu 2006, was allerdings auch bedeutet, dass die Steigerung von 2000 zu 2003 beibehalten werden konnte.

Entwicklung  
über Jahre

Die Ergebnisse der Lesekompetenz als Schwerpunktbereich in der aktuellen Erhebung 2009 sollen als erstes beschrieben werden:

PISA 2009 Lese-  
kompetenz

- Der durchschnittliche Mittelwert aller OECD-Staaten liegt bei 493 Punkten, die Standardabweichung bei 93 Punkten. Die Verschiebung des Durchschnitts lässt sich durch die veränderte Zusammensetzung der OECD-Staaten erklären.
- Bei deutschen Schülerinnen und Schülern liegt der durchschnittliche Lesekompetenzwert bei 497 Punkten und somit im Bereich des OECD-Durchschnitts.
- In Deutschland entspricht der Anteil von Schülerinnen und Schülern mit schwachen Lesekompetenzen, also jene im unteren Kompetenzniveau, 18,5 Prozent. Der OECD-Durchschnitt liegt bei 18,8 Prozent. Somit liegt Deutschland sehr genau im internationalen Mittel.
- Spitzenbereich: Der Anteil von Schülerinnen und Schülern, die die Kompetenzstufen V und VI erreichen entspricht mit 7,6 Prozent exakt dem OECD-Durchschnitt.

Die Streuung in Deutschland im Bereich der Lesekompetenz liegt bei 95 Punkten und unterscheidet sich somit kaum vom OECD-Durchschnittswert.

Streuung

Der Trend, der schon 2003 und 2006 zu beobachten war, setzt sich auch in der PISA-Erhebung 2009 fort, nämlich eine kontinuierliche Verbesserung der Lesekompetenz. Während jedoch die Verbesserungen zwischen den Jahren 2000, 2003 und 2006 nicht statistisch abzusichern waren, liegt die Lesekompetenz 2009 nunmehr mit 497 Punkten signifikant über dem im Jahr 2000 gemessenen Wert von 484 Punkten.

Trend



## Mathematik

Im Bereich Mathematik stellen sich die Ergebnisse wie folgt dar:

- Die durchschnittliche mathematische Kompetenz der OECD-Staaten liegt bei 496 Punkten.
- Deutschland liegt bei der aktuellen Erhebung mit einem mittleren Kompetenzwert von 513 über dem OECD-Mittelwert.
- Die relative Position Deutschlands hat sich im Vergleich zu seinen Nachbarstaaten gegenüber PISA 2006 verbessert.
- Zwischen den verschiedenen Bildungsgängen unterscheidet sich der Mittelwert mathematischer Kompetenz stark.
- Die mittlere mathematische Kompetenz der Schülerinnen und Schüler in Deutschland ist von PISA 2003 zu PISA 2009 signifikant angestiegen.
- Auch der Anteil der Jugendlichen unter der niedrigsten Stufe mathematischer Kompetenz ist in Deutschland von PISA 2003 zu PISA 2009 signifikant zurückgegangen.

## Streuung

Die Streuung der mathematischen Kompetenz fällt in Deutschland signifikant höher aus als im OECD-Durchschnitt, sie beträgt 98 Punkte.

## Naturwissenschaften

Als Bilanz der erreichten Ergebnisse lässt sich festhalten, dass die durchschnittliche naturwissenschaftliche Kompetenz international bei 501 Punkten liegt, Deutschland erreicht folgende Werte:

- Deutsche Schülerinnen und Schüler erreichen einen Punktwert von 520 und liegen somit signifikant über dem OECD-Mittelwert.
- Spitzenbereich: Der durchschnittliche OECD- Anteil der Schülerinnen und Schüler, die die Kompetenzstufe V und VI erreichen entspricht 8,5 Prozent, in Deutschland liegt dieser Anteil bei 12,8 Prozent.
- Der prozentuale Anteil von Jugendlichen in Deutschland auf der Kompetenzstufe I und darunter ist signifikant niedriger.
- Die mittlere naturwissenschaftliche Kompetenz der Schülerinnen und Schüler in Deutschland hat sich zwischen PISA 2006 und PISA 2009 nicht signifikant verändert.
- In den verschiedenen Bildungsgängen variieren die naturwissenschaftlichen Kompetenzen der Schülerinnen und Schüler stark. Während an den Gymnasien ein durchschnittlicher Wert von 602 Punkten erreicht wird, erreichen Schülerinnen und Schüler der Hauptschulen lediglich einen Wert von 431 Punkten.

## Streuung

Die Streuung der naturwissenschaftlichen Kompetenz beträgt in Deutschland 101 Punkte und ist signifikant höher als die durchschnittliche Streuung in den OECD-Staaten.

Für alle Bereiche zeichnet sich innerhalb der deutschen Bundesländer ein positiver Trend ab:

## Naturwissenschaften

- 2006 lagen bereits 13 Bundesländer signifikant über dem internationalen Mittelwert und nur ein Land darunter.
- Nicht geändert hat sich der weite Abstand der Bundesländer untereinander, zwischen dem besten Land (541 Punkte) und dem schlechtesten (485 Punkte) liegen 56 Punkte, dies entspricht dem Lernzuwachs von etwa zwei Jahren.
- Die besten Länder können auch im internationalen Vergleich gut mithalten.

## Lesen

- Bei der Verteilung der Kompetenzniveaus gibt es einige Länder in denen nur wenige Kinder auf oder unter der untersten Stufe liegen, in anderen Ländern ist der Anteil der Risikokinder allerdings deutlich größer.
- Auch bei der Lesekompetenz können sich die Bundesländer durchschnittlich verbessern, mehr Länder als zuvor schaffen es, oberhalb des internationalen Mittelwertes zu liegen, sind aber noch weit von den internationalen Spitzenländern entfernt.
- Der Abstand zwischen dem besten Land (Sachsen: 512 Punkte) und dem schlechtesten Land (Bremen: 474 Punkte) liegt bei 38 Punkten und hat sich damit im Vergleich zu 2003 verringert.

## Mathematik

- Da sich der allgemeine Wert für Deutschland im Bereich Mathematik von 2003 zu 2006 nur gering verändert hat, ergeben sich auch auf der Ebene der Bundesländer nur geringe Veränderungen, auch hier ist der Abstand zur internationalen Spitze ist noch groß.
- Das Land mit dem besten Ergebnis (Sachsen: 523 Punkte) liegt noch immer etwa 25 Punkte und damit international etwa ein Jahr hinter den besten Ländern.
- Die Distanz zwischen dem besten und dem schlechtesten Land hat sich zwar reduziert, liegt aber noch immer bei 45 Punkten.
- Auch die Streuung innerhalb der Länder ist im Mathematik-Bereich sehr groß.

Die Motivation der Mädchen ist im Bereich der Naturwissenschaften genauso wie ihre Freude an diesem Fächerkanon geringer als bei den Jungen. Dieses Ergebnis spiegelt sich jedoch nicht in den Kompetenzwerten wider. Allerdings sind es eher Jungen, die die oberen Kompetenzniveaus erreichen. Beim Lesen zeigt sich dagegen ein deutlicher Trend zum besseren Abschneiden der Mädchen. Dies ist sowohl im internationalen Mittel (38 Punkte besser) als auch in Deutschland (42 Punkte besser) der Fall. Es gibt aber auch Länder, wie beispielsweise die Niederlande, in denen der Unterschied mit 24 Punkten wesentlich kleiner ist. Überraschend dabei ist, dass der größte Unterschied mit 51 Punkten und damit einer halben Standardabweichung in Finnland vorliegt. Im Bereich Mathematik dagegen sind es die Jungen, die einen Vorsprung haben. Im internationalen Durchschnitt liegen sie 11 Punkte vor den Mädchen. Dabei ist der Unterschied besonders in Deutschland sehr groß, hier liegt er bei 20 Punkten, nur in Österreich (23 Punkte) und Japan (20 Punkte) ist er gleich groß oder größer.

(Deutsches PISA-Konsortium, 2001; Moschner/Kiper/Kattmann, 2003; Deutsches PISA-Konsortium, 2002; PISA-Konsortium Deutschland, 2004; PISA-Konsortium Deutschland, 2005; PISA-Konsortium Deutschland, 2007; PISA-Konsortium Deutschland, 2008)

### 3.4.2.3 IGLU (optional)

IGLU 2001 machte deutlich, dass Schülerinnen und Schüler am Ende der vierten Klasse im internationalen Vergleich gut mithalten können. Daraus kann die Schlussfolgerung gezogen werden, dass die in den Sekundarstufen durch andere Studien festgestellten Mängel nicht aus der Grundschule übernommen werden, sondern erst später entstehen.

IGLU 2001

Die Ergebnisse der deutschen Schülerinnen und Schüler kurz zusammengefasst:

- Der Mittelwert der deutschen Schülerinnen und Schüler liegt bei 539 Punkten und damit über dem internationalen Mittelwert von 500 Punkten.
- Deutschland befindet sich damit im oberen Leistungsdrittel und gleichauf mit den Staaten der Europäischen Union.
- Der Abstand zur Spitze ist groß, Schweden liegt mehr als eine Fünftel Standardabweichung entfernt.
- Zwei Drittel der Schülerinnen und Schüler erreichen Ergebnisse zwischen 472 und 606 Punkten, „die mittleren 50 Prozent der getesteten Jahrgangsstufe liegen überwiegend im Bereich des Kompetenzniveaus III“.
- Nur acht Staaten haben eine geringere Standardabweichung als Deutschland (67 Punkte), somit ist Deutschland einer der wenigen Staaten mit geringer Streuung.
- In Deutschland liegen nur 1,3 Prozent der Schülerinnen und Schüler unter Kompetenzniveau I, international liegt dieser Wert bei 11,6 Prozent, in Vergleichsgruppe (VG) 2 (siehe Kapitel 3.2.1.2) bei 2,9 Prozent.
- Das höchste Kompetenzniveau (IV) wird in Deutschland von 18,1 Prozent der Schülerinnen und Schüler erreicht (international: 13,7 Prozent, VG 2: 18 Prozent).
- Einige Staaten (z. B. Frankreich), die in PISA deutlich besser abgeschnitten haben als Deutschland, liegen bei IGLU deutlich hinter Deutschland.

Kompetenz-  
stufen

In IGLU-E wurden 2001 die Bereiche Naturwissenschaften, Mathematik und Orthografie untersucht. Die Ergebnisse lauten:

Ergänzungs-  
studie

Naturwissenschaften: Ergebnisse können mit denen von TIMSS verglichen werden, für die Schülerinnen und Schüler ergibt sich ein Mittelwert von 560 Punkten, dieser muss im Vergleich zu 524 Punkten als Mittelwert von TIMSS gesehen werden (TIMSS: internationaler Mittelwert 500, aber 3. und 4. Klasse, 4. Klasse 524 Punkte, 3. Klasse 473 Punkte)

Natur-  
wissenschaften

- Deutschland liegt also über dem internationalen Mittelwert und im Vergleich mit den Ergebnissen von TIMSS im oberen Drittel.
- Mathematik: Vergleiche mit den Ergebnissen von TIMSS möglich, der Mittelwert der deutschen Schülerinnen und Schüler liegt bei 545 Punkten (internationaler Mittelwert: 529 Punkte), Deutschland liegt auch hier deutlich oberhalb des internationalen Mittelwerts.
- Ergebnisse der besten Staaten werden nicht erreicht.
- Orthografie: 45 Wörter diktirt, davon durchschnittlich 25,6 korrekt aufgeschrieben
- Standardabweichung bei 9,0 Wörtern
- Nur sehr wenige Kinder (3 von 2.951) schreiben alle Wörter richtig, die besten fünf Prozent der Schülerinnen und Schüler schreiben 40 Wörter richtig, die unteren 15 Prozent schreiben mehr als 30 Wörter falsch.

Mathematik

Orthografie

Auch in IGLU 2001 lässt sich ein Vorsprung der Mädchen beim Lesen erkennen. Allerdings ist dieser im internationalen Vergleich nicht so groß, da er in Deutschland bei 13 Punkten, im internationalen Vergleich dagegen bei 20 Punkten liegt.

Erkenntnisse  
aus Frage-  
bögen

Zudem sollen einige Aspekte betrachtet werden, die durch die zusätzlichen Erhebungen anhand von Fragebögen erkennbar wurden und Erklärungsmöglichkeiten für die Ergebnisse bieten:

- Die Klassengröße in Deutschland liegt im internationalen Durchschnitt, allerdings gibt es weniger zusätzliche Lehrer, die in den großen Klassen eingesetzt werden.
- Die Lehrkräfte an deutschen Grundschulen sind im Vergleich älter, sie erhalten ein besonders hohes Gehalt.
- In vielen Staaten der Vergleichsgruppe 2 stehen den Schülerinnen und Schülern mehr Computer zur Verfügung, sodass dort der Computer häufiger als Arbeitsmittel eingesetzt werden kann.
- Schülerinnen und Schüler in anderen Staaten verfügen bereits bei der Einschulung über Lesevorkenntnisse, da diese dort im Rahmen des Kindergartens ausgebildet werden, in Deutschland ist dies nicht der Fall.
- Deutsche Schulleiterinnen und Schulleiter sind vor allem Lehrkräfte, dies ist in vielen anderen Staaten anders, hier sind die Schulleiterinnen und Schulleiter für die Qualität der Schule und des Unterrichts zuständig und unterrichten nur sehr wenig.
- In vielen Staaten werden Lehrerinnen und Lehrer dazu angehalten, untereinander zu kooperieren, dies wird durch feste Zeiten und Absprachen unterstützt, in Deutschland basiert eine Zusammenarbeit der Lehrkräfte auf Einzelinitiativen. Differenziert wird in Deutschland vor allem, um schwache Schülerinnen und Schüler zu unterstützen, dies ist in den meisten Staaten so, dazu erhalten diese Schülerinnen und Schüler in Deutschland mehr Bearbeitungszeit für die gleichen Aufgaben, dieses Verfahren wird ebenfalls von mehreren Staaten angewendet.

Kooperation  
von Lehr-  
kräften

IGLU 2006

Bei IGLU 2006 erreicht Deutschland einen Mittelwert von 548 Punkten, liegt damit weit über dem internationalen Mittelwert (506 Punkte) und auch über den Werten der Vergleichsgruppen. Mit diesem Ergebnis liegt Deutschland im oberen Viertel aller teilnehmenden Staaten. Innerhalb der Europäischen Union gibt es kein Land, das besser abschneidet als Deutschland. Die Standardabweichung der deutschen Schülerinnen und Schüler liegt bei 67 Punkten und damit weit unter dem internationalen Durchschnitt von 103 Punkten. In der deutschen Stichprobe gibt es nur wenige Schülerinnen und Schüler, die als Risikokinder gelten, nur zwei Staaten erreichen einen geringeren Anteil. Allerdings ist der Anteil der Spitzenleser (sie erreichen das höchste Kompetenzniveau) nicht befriedigend, da er nur bei 10,8 Prozent liegt.

Kompetenz-  
stufen

Im intranationalen Vergleich zeigen sich folgende Ergebnisse:

- Lesekompetenz: große Unterschiede, Thüringen mit einem Mittelwert von 564 Punkten weit vor Bremen mit 522 Punkten. Thüringen erreicht damit genauso viele Punkte wie Hongkong und liegt weit vorne an der Spitze.
- Die Streuung der Bundesländer ist sehr unterschiedlich und reicht von 59 Punkten (Thüringen, Sachsen-Anhalt, Rheinland-Pfalz) bis zu 76 Punkten (Berlin).
- Die Verteilung auf die Kompetenzniveaus fällt in den Bundesländern sehr unterschiedlich aus, so gibt es Länder, die wenige Risikokinder und viele Spitzenleser haben (Thüringen: 0,4 Prozent auf Stufe I, 14,9 Prozent auf Stufe V) und Länder, die viele Risikokinder und wenige Spitzenleser haben (Berlin: 6,7 Prozent auf Stufe I, 8,6 Prozent auf Stufe V).
- Unterschied zwischen Mädchen und Jungen: sehr unterschiedliche Ergebnisse. Der Vorsprung der Mädchen reicht von -1 Punkt in Berlin (also ein Vorsprung der Jungen) bis zu +18 Punkten (Vorsprung der Mädchen) in Sachsen-Anhalt.

Bundeslän-  
der im Ver-  
gleich

Ergänzende  
Aspekte

Auch 2006 wurden einige ergänzende Aspekte festgestellt. Ebenso wie aus der Erhebung von 2001 ergeben sich Hinweise darauf, dass sich die Vorschulzeit auszahlt: Besuchen Kinder eine vorschulische Einrichtung, so verbessert sich beispielsweise ihre Lesekompetenz. Dieses Ergebnis konnte sowohl international als auch für Deutschland erkannt werden. Zudem wurde deutlich, dass deutsche Schülerinnen und Schüler aus bildungsnahe Elternhäusern einen deutlichen Vorsprung von 67 Punkten vor Schülerinnen und Schülern aus bildungsfernen Elternhäusern haben. Dies ist ein Aspekt, der auch durch PISA festgestellt wurde. Der Vorsprung ist in Deutschland deutlich größer als das internationale Mittel von 57 Punkten. Auch in IGLU 2006 zeigte sich, dass Mädchen besser lesen als Jungen. Allerdings ist Deutschland dabei einer der sehr wenigen Staaten, in denen die Differenz mit 7 Punkten (internationales Mittel: 13 Punkte) sehr gering ausfällt. Zudem wurde gezeigt, dass Schülerinnen und Schüler mit Migrationshintergrund im internationalen Mittel schlechter abschneiden als Schülerinnen und Schü-

Bildungsnahe  
/ bildungs-  
ferne Eltern-  
häuser

ler ohne Migrationshintergrund. Dieser Unterschied konnte allerdings in Deutschland von 2001 zu 2006 um 7 Punkte verringert werden.

Abschließend kann festgehalten werden, dass Deutschland zu den elf teilnehmenden Staaten gehört, die 2006 eine signifikant bessere Leistung erreichen als 2001. In Deutschland wird eine Steigerung von 9 Punkten erreicht. Die größte Verbesserung kann in Russland (+37), Hongkong (+36) und Singapur (+30) festgestellt werden. Dabei gibt es auch Staaten, die sich innerhalb der fünf Jahre extrem verschlechtert haben. Dazu gehören Kuwait (-66), Marokko (-27) und Rumänien (-22). Zudem kann festgestellt werden, dass Deutschland näher an die besser abscheidenden Staaten herankommt und sich von den schlechter Abscheidenden absetzen kann.

(Bos u.a., 2003; Bos u.a., 2007; Bos u.a., 2008)

#### 3.4.2.4 DESI (optional)

Die DESI-Ergebnisse von Englisch als erster Fremdsprache orientieren sich am Gemeinsamen europäischen Referenzrahmen für Sprachen (GeR), der bereits in Studienbrief-Teil ‚Vergleichsarbeiten‘ thematisiert wurde. Die Ergebnisse im Bereich der mündlichen Sprechfähigkeit sagen Folgendes aus:

- Zwei Drittel der deutschen Schülerinnen und Schüler erreichen am Ende der 9. Klasse das Niveau A2<sup>19</sup> (Erwartungshorizont für den Hauptschulabschluss), Schülerinnen und Schüler, die sich auf diesem Niveau befinden, können sich im Alltag verständlich machen.
- Ein weiteres Drittel der Schülerinnen und Schüler erreicht bereits das Niveau B1 (Erwartungshorizont für den mittleren Schulabschluss am Ende der 10. Klasse).
- Neun Prozent der Schülerinnen und Schüler können sich auf einem noch höheren Niveau verständlich machen.

Ähnliche Ergebnisse zeigen sich auch für die anderen getesteten Bereiche. Es kann zusammengefasst werden, dass der Englischunterricht, vor allem an Gymnasien, dazu führt, dass sich eine starke Leistungsspitze von 10 bis 15 Prozent bildet, die Kompetenzen besitzt, die weit über das Geforderte hinaus gehen.

Aber auch negative Aspekte können durch DESI sichtbar werden. So haben Hauptschulen, integrierte Gesamtschulen und Schulen mit mehreren Bildungsgängen deutliche Defizite. Dies zeigt sich am Beispiel der Hauptschule darin, dass nur etwa ein Drittel der Schülerinnen und Schüler das von den Bildungsstandards aufgestellte Regelziel in Englisch erreicht.

Die Ergebnisse für Deutsch lauten zusammengefasst:

- Zwei Drittel der Schülerinnen und Schüler sind am Ende der 9. Klasse in der Lage, Briefe verständlich zu formulieren, 13 Prozent der Schülerinnen und Schüler können sich dabei stilsicher und abwechslungsreich ausdrücken.
- Drei Viertel der Schülerinnen und Schüler erkennen klare grammatische Verstöße, schwierige Verstöße werden von einem Drittel erkannt.
- Lesekompetenz, bezogen auf literarische und Sachtexte: Fast alle Schülerinnen und Schüler erreichen mindestens das unterste Niveau.
- Ein Drittel kann zielgerichtet lesen und Informationslücken schließen.
- Übergeordnete Textstrukturen können von jedem Sechsten erkannt und mit eigenem Wissen verbunden werden.
- Etwa die Hälfte der Schülerinnen und Schüler an Haupt- und Gesamtschulen schreibt nicht angemessene Texte und erkennt keine einfachen grammatischen Fehler.
- Gymnasium: Ein Drittel der Schülerinnen und Schüler kann stilsichere und nahezu fehlerfreie Texte schreiben.

Besondere Erkenntnisse ergeben sich bei DESI daraus, dass die Schülerinnen und Schüler zu zwei Zeitpunkten getestet wurden und sich daher Veränderungen über ein Schuljahr aufzeigen lassen. Diese liegen vor allem im Englischen, aber auch in der Sprachbewusstheit des Deutschen und werden zusammengefasst in Tabelle 15 dargestellt.

Schülerinnen und Schüler mit Migrationshintergrund

Entwicklung 2001 - 2006

Mündliche Sprechfähigkeit

Negative Aspekte

Deutsch

Schreibkompetenz

Lesekompetenz

Veränderungen über ein Schuljahr

<sup>19</sup> Der Gemeinsame europäische Referenzrahmen für Sprachen ist eingeteilt in die Niveaus A1, A2, B1, B2, C1 und C2, wobei in DESI nur von A1 bis C1 getestet wurde.

Kompetenzbereich	Veränderung
Hörverstehen/Englisch	Steigerung um 27 Punkte
Umgang mit englischsprachigen Texten	Steigerung um 23 Punkte
Umgang mit Grammatik und Sprachstil/Deutsch	Zuwachs um 35 Punkte
Lesekompetenz	Kein messbarer Anstieg
Schreiben	Inhaltliche Qualität der Texte geringfügig verbessert

Tabelle 15: Veränderungen in unterschiedlichen Kompetenzbereichen über ein Schuljahr bei DESI

Zudem wird erkennbar, dass der Deutschunterricht in allen Schulformen gleich wirksam ist, da sich die Unterschiede zwischen den Schulformen innerhalb eines Schuljahres nicht verändern. Dies ist im Englischen anders, da hier vor allem die Gymnasiasten einen großen Kompetenzzuwachs zeigen.

Mädchen /  
Jungen

Deutlich wird durch DESI, dass die Mädchen bei sprachbezogenen Aufgaben deutlich vor den Jungen liegen. Bei DESI liegt der Vorsprung der Mädchen in Deutsch bei 41 Punkten und in Englisch bei 31 Punkten. Schülerinnen und Schüler mit nicht-deutscher Erstsprache liegen deutlich hinter denen mit Deutsch als Erstsprache zurück. Schülerinnen und Schüler, die mehrsprachig aufwachsen, liegen dagegen weniger zurück. Schülerinnen und Schüler mit nicht-deutscher Erstsprache fällt das Erlernen von Englisch leichter. Dies trifft besonders auf die Schülerinnen und Schüler zu, die mehrsprachig aufgewachsen sind.

Unterrichts-  
qualität

Zudem wurden auch Aspekte der Unterrichtsqualität untersucht. Der wichtigste Befund dabei ist, dass „die Lehrkraft im Durchschnitt doppelt so viel spricht wie alle Schüler zusammen“. Die Schülerinnen und Schüler haben also im Unterricht nur sehr selten die Möglichkeit, ihre Sprachkompetenzen auszuprobieren und zu erweitern.

(Klieme, 2006; DESI-Konsortium, 2008)

### 3.4.3 Verwendung der Ergebnisse

Wichtig bei der Darstellung und Interpretation der Ergebnisse ist, dass die Studien des Systemmonitoring vor allem als regelmäßiges Instrument gedacht sind und daher nicht nur die Ergebnisse zu einem Zeitpunkt, sondern vielmehr der Vergleich von mindestens zwei aufeinanderfolgenden Testzeitpunkten betrachtet werden muss (Wolter, 2008). Eine solche Interpretation wird in Zukunft dadurch vereinfacht, dass die Studien in einem regelmäßigen Rhythmus durchgeführt werden.

Festzuhalten ist, dass die Studien nicht zu einem rein wissenschaftlichen Zweck durchgeführt werden. Ihr Ziel ist es (vgl. Kapitel 1.2), „professionelles pädagogisches Handeln in der Unterrichtspraxis, in Schuladministration und Bildungsverwaltung zu unterstützen“. Daher werden die Ergebnisse so aufbereitet, dass sie von einer breiten Mehrheit wahrgenommen und kritisch hinterfragt werden können. Nur wenn die Ergebnisse verständlich sind, können sie für das eigene Handeln berücksichtigt und übernommen werden (Klieme/Baumert, 2001). Veränderungen im Unterrichtsalltag ergeben sich einerseits aus den reflektierten Handlungen der Lehrkräfte und andererseits aus den Impulsen der Bildungspolitik und der Bildungsadministration.

Verwendung  
der Studien  
im Schulalltag

In Studien des Systemmonitorings werden Ergebnisse für die Schülerpopulation eines Landes geschätzt. Diese müssen nicht mit den Kompetenzen der Schülerinnen und Schüler an der eigenen Schule und in der eigenen Klasse übereinstimmen. Allerdings ist es in einigen Fällen möglich, Instrumente des Systemmonitorings auch in einzelnen Schulen einzusetzen, wenn das Instrument einen Aspekt untersucht, der für die Schule von besonderem Interesse ist. Die Daten der Studie können dann als ‚Normierungsstudie‘ benutzt und die Ergebnisse der Schule beziehungsweise Klasse können im Vergleich zu anderen Schulen verortet werden (Klieme u.a., 2007). Ob eine solche Verwendung der Ergebnisse möglich ist, muss für jede einzelne Studie geprüft werden.

Rückmeldung an  
teilnehmende  
Schulen

Eine weitere Möglichkeit, die Ergebnisse der Studien für die eigene Arbeit zu nutzen, besteht, wenn die eigene Schule beziehungsweise Klasse an der Studie teilgenommen hat. Dann werden in einigen Studien die Ergebnisse der Schule zurückgemeldet. Dabei müssen die Rückmeldungen so gestaltet werden, dass die Schule die Informationen aufgreifen und reflektieren kann. Dabei darf die Erwartung an eine solche Rückmeldung aber nicht zu groß sein. Zudem müssen die Ziele und Probleme der jeweiligen Schule beachtet werden (Klieme/Baumert, 2001).

Umgang mit  
Daten

Wichtig ist es auch zu beachten, dass Schulen mittlerweile eine Unmenge von Daten zur Verfügung gestellt bekommen und mit diesen auch umgehen können müssen. Dazu müssen auch die Kompetenzen innerhalb der Schule vorhanden sein, sodass die Ergebnisse interpretiert werden können und Begrifflichkeiten nicht unklar bleiben.

Zudem kann es von Schule zu Schule unterschiedlich sein, wie ein bestimmter Wert interpretiert wird. Dies hängt mit den unterschiedlichen Vorstellungen und Modellen von Schulen zusammen (Rolff, 2008).

Ein Problem besteht darin, dass die Erfassung von Daten im Rahmen des Systemmonitorings in den Schulen als Intervention empfunden wird. Dieses Spannungsfeld entsteht immer dann, wenn wissenschaftliche Ergebnisse innerhalb der Praxis, in diesem Fall der politischen und pädagogischen, benutzt werden sollen. „Die Handelnden in der erstgenannten Welt wissen mehr als sie können, die Handelnden in der letztgenannten können mehr als sie wissen“ (Rolff, 2008). Daher genügt es nicht, wenn die Ergebnisse in ihrer wissenschaftlichen Art und Weise an die Schulen zurückgegeben werden, sie müssen aufbereitet werden. So kann es gelingen, dass die Ergebnisse nicht als Intervention sondern als „Handlungsimpulse“ verstanden werden können und eine positive Entwicklung stattfinden kann (Rolff, 2008). Dies sollte das Ziel aller am Systemmonitoring beteiligten Personen sein.

Systemmonitoring = Intervention?

### 3.4.4 Weiterführende Literatur

Ausführliche Informationen über die Ergebnisse der einzelnen Studien bieten die Berichte der Studien. Die Literaturangaben können den einzelnen Kapiteln entnommen werden.

### 3.4.5 Verständnis-Aufgaben und Diskussionspunkte

1. *Betrachten Sie die Tabelle der Kompetenzniveaus für TIMSS 1995 und erstellen Sie Interpretationen für die anderen vier Staaten.*
2. *Erstellen Sie eine Tabelle mit den Ergebnissen aller drei PISA-Erhebungen (zu einem selbst gewählten Kompetenzbereich) und machen Sie Entwicklungen über die Jahre deutlich.*
3. *Für IGLU 2001 wird angegeben, welche Erklärungsmöglichkeiten sich durch zusätzliche Erhebungen (beispielsweise nach der Klassengröße) für die Ergebnisse ergeben. Überlegen Sie, für welche Ergebnisse die Erklärungsmöglichkeiten passen könnten.*
4. *Überlegen Sie, wie es dazu kommen kann, dass sich Staaten im Verlauf von einigen wenigen Jahren stark verbessern oder verschlechtern (Beispiel IGLU 2001 zu 2006).*



### 3.5 Praktische Implikationen

Kapitel 3.5 beschäftigt sich abschließend mit den praktischen Implikationen. Hier wird betrachtet, welche Auswirkungen sich durch die Systemmonitoring-Studien ergeben, bezogen auf die Bildungspolitik allgemein, vor allem aber auf den Schulalltag.

In diesem Kapitel werden folgende Fragen beantwortet:

- *Wie werden die Studien von den Beteiligten rezipiert?*
- *Welche politischen Entscheidungen wurden auf Grund der Ergebnisse getroffen?*
- *Welche Auswirkungen auf die Schulpraxis ziehen die Studien-Ergebnisse nach sich?*

Wie bereits in den vorangegangenen Kapiteln beschrieben, führte TIMSS 1995 dazu, dass vermehrt Aktivitäten im Bildungswesen unternommen wurden, und zwar in drei verschiedenen Bereichen:

1. Es wurden Reformmaßnahmen für den mathematisch-naturwissenschaftlichen Unterricht beschlossen.
2. Im Bildungswesen erfolgte eine Wende mit einem Fokus auf empirische Forschung.
3. Systematisches Bildungsmonitoring wurde in seiner bereits beschriebenen Form eingeführt (Klime/Baumert, 2001).

Spätestens PISA 2000 hat dazu geführt, dass das deutsche Bildungssystem grundlegend in Frage gestellt wurde. Was daraus folgte und aus späteren Untersuchungen noch folgen wird, sind „Veränderungen im Schulsystem, in didaktisch-methodischen Orientierungen und in der Lehrerbildung“ (Eichler, 2003). Ziel all dieser Veränderungen soll es sein, „die Qualität der Bildungssysteme zu steigern und zu sichern“ (Stanat, 2008). Dieses Kapitel soll daher zeigen, welche Folgen sich aus der Einführung eines systematischen Bildungsmonitorings in Deutschland bisher ergeben haben. Dazu sollen mehrere Aspekte beschrieben werden, die als direkte Konsequenz aus den durch die Studien gewonnenen Informationen angesehen werden können. Zuerst werden allgemein bildungspolitische Veränderungen, anschließend dann ganz konkrete Entwicklungen in der Schulpraxis betrachtet.

Zunächst soll dargestellt werden, wie die Ergebnisse von den unterschiedlichen Akteuren im Bildungsbereich aufgenommen werden. Bei der Rezeption der Ergebnisse kam es auch zu Fehlinterpretationen, sodass die Reaktionen auf die Ergebnisse nicht immer sachgemäß waren. So wurden beispielsweise die Rangskalen mehr betrachtet als die Aussagekraft der Mittelwerte und Standardabweichungen (vgl. dazu Kapitel 3.4.2.1 bis 3.4.2.4). Einige Projekte wurden allerdings auch überhastet und ohne kritische Reflexion umgesetzt, da viele beteiligte Personen davon ausgingen, dass möglichst schnell gehandelt werden muss. Beispielsweise ist dies bei der Einführung zentraler Abiturprüfungen geschehen, da sich nicht aus den Ergebnissen ablesen lässt, dass dies zu einer zwangsläufigen Verbesserung führt (Moschner/Kiper/Kattmann, 2003).

Beachtet werden muss, dass die Übersetzung der Ergebnisse in tatsächliche Veränderungsvorschläge alles andere als trivial ist. Eine Schwierigkeit liegt darin, dass aus den Ergebnissen nicht direkt entnommen werden kann, an welchen Stellen Veränderungen ansetzen müssen. Zudem ist es schwer zu bestimmen, unter welchen Bedingungen sich die vorgeschlagenen Veränderungen umsetzen lassen, wie sie von der Praxis aufgenommen und welchen Effekt sie tatsächlich haben werden. Eine Gefahr besteht vor allem darin, dass die Veränderungen auch unerwünschte Nebeneffekte zur Folge haben können, vor allem dadurch, dass nur noch die leicht überprüfbaren Ziele des Bildungssystems verfolgt werden könnten (Stanat, 2008). Darüber hinaus lassen sich Fehl- oder Überinterpretationen und daraus folgende, unreflektierte Veränderungen nicht ausschließen. Dies geschieht vor allem dann, wenn aus den empirischen Daten direkte Handlungskonsequenzen für die Schulpraxis abgeleitet werden (Stanat, 2008).

#### Beispiel für Fehl- beziehungsweise Überinterpretationen

PISA-Befund: Die sozialen Disparitäten sind in Deutschland besonders ausgeprägt.

Begründung: Dies liegt daran, dass die Schülerinnen und Schüler schon früh auf die verschiedenen Schulformen verteilt werden.

Forderung: Reform der Schulstruktur

ABER: Kein Beleg in den PISA-Befunden dafür, dass die Ergebnisse durch die frühe Verteilung bedingt sind!

(nach: Stanat, 2008)

Fehler bei der  
Rezeption

Übersetzung der  
Ergebnisse



Um solche Fehlinterpretationen zu vermeiden, ist es ratsam, bei der Interpretation der Ergebnisse und der Ableitung von Handlungen auf mehrere Studien zurückzugreifen, da so mehr Informationen gebündelt betrachtet werden können (Stanat, 2008). Für das oben genannte Beispiel könnten beispielsweise weitere Studien zu den Themen ‚soziale Disparitäten‘ und ‚Schulstruktur‘ zu Rate gezogen werden.

Die Resultate der regelmäßig durchgeführten Studien zeigen, dass sich die Ergebnisse deutscher Schülerinnen und Schüler langsam verbessern. Das lässt darauf schließen, dass die bisherigen Veränderungen beginnen zu wirken. Ziel der Bildungspolitik muss es sein, diese positive Entwicklung weiter voran zu treiben und sich nicht auf den aktuellen Verbesserungen ‚auszurufen‘, denn neben den positiven Entwicklungen zeigen die Ergebnisse der aktuellen Studien auch, dass noch immer viele Staaten deutlich bessere Ergebnisse erreichen. Hier muss vor allem erforscht werden, warum sich in anderen Staaten bessere Ergebnisse zeigen. Dies ist allerdings schwierig, da, wie bereits beschrieben, nicht festgestellt werden kann, welcher Aspekt das gute Abschneiden bedingt. So ist in Hinblick auf das obige Beispiel festzustellen, dass es sowohl Staaten ohne als auch mit gegliedertem Schulsystem gibt, die in den Studien schlecht abschneiden (Bos/Postlethwaite, 2001).

### 3.5.1 Rezeption der Studien (optional)

Die Ergebnisse der Systemmonitoring-Studien werden sehr unterschiedlich aufgefasst. Dabei ist leider auffällig, dass die Rezeption manchmal stark verkürzt oder sehr selektiv durchgeführt wird. In einigen Fällen werden den Studien auch Argumente entnommen, die durch die Studie selbst nicht bestätigt werden konnten (Kohler, 2005).

In diesem Kapitel soll dargestellt werden, wie die Ergebnisse der Studien von den verschiedenen Akteuren der Bildung aufgenommen werden. Weinert (2001c) verwandelt diese ernste Szene in ein ironisches Szenario und gibt ein Beispiel, wie ein Gymnasiallehrer auf die Ergebnisse von TIMSS reagiert haben könnte:

„Was mögen seine Gedanken, seine Gefühle und seine Erwartungen gewesen sein? Abschiebung jeglicher Verantwortung auf die Politik, auf die Schuladministration oder auf das mehr oder minder anonym definierte deutsche Bildungswesen? Impulsive Infragestellung der berichteten empirischen Befunde; Bedenken über die Vergleichbarkeit von Leistungswerten aus unterschiedlichen Ländern; Grimm über die Nichtberücksichtigung nachteiliger deutscher Bildungsfaktoren? Waren es persönliche Betroffenheit, Ärger, vielleicht sogar Zorn über eine neue Störung des schwierigen Dialogs zwischen Schule und Öffentlichkeit, zwischen überforderten einzelnen Lehrern und besorgten Eltern?“

Ähnliche Vermutungen lassen sich auch für die anderen Beteiligten anstellen. Dabei wird schnell nachvollziehbar, dass vor allem Bildungspolitikern dazu neigen, schnelle Lösungen finden zu wollen, um sich nicht mehr entsprechenden Fragen stellen zu müssen. Sinnvolle Reaktionen benötigen aber einen bestimmten Zeitraum um sich zu entwickeln, und bessere Ergebnisse können sich nicht von einem Tag auf den anderen einstellen (Weinert, 2001c).

Kohler (2005) hat eine Studie durchgeführt, um die Rezeption der TIMSS-Ergebnisse von TIMSS der unterschiedlichen Akteure zu untersuchen. Dabei wurden die folgenden drei Personengruppen befragt:

Studie zur  
Rezeption

- Lehrerinnen und Lehrer
- Eltern
- Schuladministration

Einige besonders interessante und hilfreiche Ergebnisse dieser Studie sollen hier kurz vorgestellt werden. Informationen zu Anlage und Durchführung dieser Studie können der Literatur entnommen werden.

Durch Kohlers Untersuchung wurde deutlich, dass sich die Lehrerinnen und Lehrer über die Studie (in diesem Fall: TIMSS) für nicht gut informiert hielten und ihr Vorwissen darüber eher gering war. Die Einstellungen der Lehrerinnen und Lehrer zu TIMSS waren sehr unterschiedlich, sie reichten von „Was bringt so eine Studie letztendlich?“ bis zu „Man wurde über eigene Meinung durch die Studie bestätigt.“ Zudem schätzten Lehrerinnen und Lehrer, egal welcher Schulform, Fächer und Geschlecht, die Bedeutung von TIMSS gleich ein. Allerdings fanden Lehrerinnen und Lehrer mit Leitungsfunktion die Studie bedeutsamer. Auf die Frage danach, warum die deutschen Schülerinnen und Schüler nicht gut abgeschnitten haben, suchten Lehrerinnen und Lehrer die Antworten vermehrt außerhalb und nicht innerhalb der Schule. Allerdings sahen viele Lehrerinnen und Lehrer die Notwendigkeit, auf Grund der Ergebnisse etwas an ihrem Unterricht zu verändern. An die Bildungspolitik stellten sie Forderungen nach kleineren Klassen, reformierten Lehrplänen, mehr Förderunterricht und Fortbildungsmaßnahmen (Kohler, 2005).

Lehrerinnen  
und Lehrer

Eltern

Auf der Seite der Eltern kann ganz allgemein davon ausgegangen werden, dass sich ihr Interesse an der Bildung ihres Kindes vergrößert hat. Sie haben bestimmte Ansprüche an die Schule und möchten mitentscheiden, wie der weitere Bildungsweg ihres Kindes gestaltet wird. Problematisch kann dabei sein, dass die Eltern jeweils nur ihr eigenes Kind im Blick haben und daher Vorstellungen entwickeln, die nicht für eine Förderung aller Schülerinnen und Schüler hilfreich sind (Kohler, 2005). Es ist wichtig, dass Eltern eine positive Einstellung den Systemmonitoring-Studien gegenüber haben, da sie im Prozess der Qualitätsentwicklung von Schulen eine wichtige Rolle spielen.

Die in der Studie untersuchten Eltern hielten sich für sehr wenig über die Studie informiert und verfügten nur über ein sehr geringes Vorwissen, deutlich geringer als das der Lehrerinnen und Lehrer. Generell fielen die Äußerungen der Eltern über das Abschneiden der deutschen Schülerinnen und Schüler wenig homogen aus. In einem mittleren Maß fanden die Eltern die Studie bedeutsam, dies kann unter anderem damit begründet werden, dass die Eltern sich mehr für das Abschneiden des eigenen Kindes interessieren. Gegenüber den Studien des Systemmonitorings waren die Eltern allgemein sehr positiv eingestellt, allerdings waren sie, ähnlich wie die Lehrerinnen und Lehrer, skeptisch, ob sinnvolle internationale Vergleiche möglich sind. Auch die Eltern suchten die Begründungen für das schlechte Abschneiden extern und nicht bei sich. Dabei wurde besonders eine generelle Unzufriedenheit mit der Bildungspolitik deutlich. Anders als die Lehrerinnen und Lehrer sahen die Eltern allerdings nur in sehr geringem Maße eine Bedeutung der Ergebnisse für die eigene Person und die eigenen Kinder (Kohler, 2005).

Unter dem Begriff der Schuladministration werden bei Kohler (2005) Beamtinnen und Beamte der Schulaufsicht verstanden. Sie haben einen großen Erfahrungshintergrund, kennen die praktische Arbeit der Lehrerinnen und Lehrer und sind innerhalb landesweiter Reformen beteiligt.

Schuladministration

Bei den Ergebnissen der Studie von Kohler finden sich viele Übereinstimmungen der Schuladministration mit den Lehrerinnen und Lehrern und den Eltern. Die Unterschiede liegen darin, dass die Schuladministration die Bedeutung der Studie als groß einschätzte und ihren Nutzen deutlich anerkannte. Zudem wünschte sie sich mehr Informationen. Dies kann daran liegen, dass Informationen über die Leistungen des gesamten Schulsystems für die Bildungsadministration auf den ersten Blick von wesentlich größerer Bedeutung sind als für die Lehrerinnen und Lehrer und die Eltern. Bei der Frage nach Handlungsmöglichkeiten hat die Schuladministration vor allem die Kompetenzen der Lehrerinnen und Lehrer im Blick. Hier wurde eine verbesserte Aus- und Fortbildung der Lehrerinnen und Lehrer gefordert. Die Aussage, die Ergebnisse seien dadurch begründet, dass zu wenig Geld für die Schulen zur Verfügung gestellt wurde, fand in der Bildungsadministration keine Unterstützung. Ebenfalls war hier kaum die Meinung vertreten, dass die schulischen Rahmenbedingungen verändert werden müssten, ein Aspekt, der bei den Eltern und vor allem bei den Lehrerinnen und Lehrern von großer Bedeutung war.

Somit wird deutlich, dass die Ergebnisse der Studien unterschiedlich aufgefasst werden können. Daher ist es für alle Beteiligten von großer Bedeutung, dass sie ihre ersten Einschätzungen überprüfen und ihre Meinungen auf einer breiten Basis an Informationen formulieren.

### 3.5.2 Entscheidungen auf Grund der Ergebnisse

Bereits in Kapitel 3.1 wurde beschrieben, dass die KMK nach den schlechten Ergebnissen von TIMSS handeln musste und dazu die ‚Konstanzer Beschlüsse‘ verfasste. Darin wurde eine regelmäßige Teilnahme an weiteren internationalen Studien beschlossen, um herauszufinden, welche Kompetenzniveaus von den deutschen Schülerinnen und Schülern erreicht werden können. Da daraufhin die deutschen Schülerinnen und Schüler auch in PISA 2000 schlecht abschnitten, wurden weitere Formen der Überprüfung wie beispielsweise Vergleichsarbeiten (vgl. Studienbrief-Teil Vergleichsarbeiten) eingeführt.

Im Jahr 2004 wurde von der KMK das ‚Institut zur Qualitätsentwicklung im Bildungswesen – Wissenschaftliche Einrichtung der Länder an der Humboldt-Universität zu Berlin‘ (IQB) gegründet. Eine Aufgabe des Instituts ist es, die notwendigen empirischen Testverfahren vorzubereiten (KMK, 2006). Außerdem ist es Aufgabe des Instituts, Referenzaufgaben zu entwickeln und Handbücher zu verfassen, die für die Umsetzung und Überprüfung der Bildungsstandards (vgl. Kapitel 3.5.3.1) eingesetzt werden (PISA-Konsortium Deutschland, 2007). Auf Seiten der KMK wurde zudem die Arbeit an Bildungsstandards vorangetrieben und die zentralen Abiturprüfungen entwickelt (Köller, 2008). Da sich diese Entwicklungen direkt auf den Schulalltag auswirken, werden sie genauer in Kapitel 3.5.3 beschrieben.

IQB

### 3.5.2.1 Sieben Handlungsfelder der KMK nach PISA 2000

Am 04.12.2001 wurden nicht nur die ersten Ergebnisse von PISA 2000 vorgestellt, sondern auch ein Handlungskatalog der KMK. Darin wurden sieben Handlungsfelder beschrieben, in denen in naher Zukunft Maßnahmen unternommen werden sollten, um auf die schlechten Ergebnisse der deutschen Schülerinnen und Schüler zu reagieren (Tillmann, 2008):

1. Verbesserung der Sprachkompetenzen in verschiedenen Bereichen
2. Bessere Verzahnung von Vor- und Grundschule; frühere Einschulung
3. Verbesserung der Grundschulbildung
4. Bessere Förderung bildungsbenachteiligter Kinder
5. Qualitätssicherung durch verbindliche Standards und Evaluation
6. Stärkung der diagnostischen und methodischen Kompetenzen der Lehrkräfte
7. Ausbau schulischer und außerschulischer Ganztagsangebote

Viele der Handlungsfelder sind komplex und beinhalten mehrere Ziele und mögliche Bereiche der Veränderung. Wie sich diese Handlungsfelder in der Praxis niedergeschlagen haben, wird Kapitel 3.5.3 zeigen. Hier werden Veränderungen in der Praxis beschrieben und mit dem jeweiligen Handlungsfeld verknüpft.

Die Beschlussfassung der Handlungsfelder führte dazu, dass die bis zu diesem Zeitpunkt teilweise sehr unterschiedlichen Aktivitäten der sechzehn Bundesländer gebündelt wurden. Bereits vor der Beschlussfassung führten einige der Bundesländer Maßnahmen in einem oder mehreren Handlungsfeldern durch. Es wird davon ausgegangen, dass diese Länder versucht haben, ihre Maßnahmen auch in der gesamtdeutschen Beschlussfassung unterzubringen, um so ihre eigene Politik abzusichern und zu legitimieren. Somit wird deutlich, dass die Handlungsfelder nicht vollkommen neue Ideen widerspiegeln, sondern bereits verfolgte Ideen zusammenfassten und für ganz Deutschland thematisierten. Das bedeutet allerdings auch, dass einige Maßnahmen unabhängig von PISA entwickelt, nach der Beschlussfassung der KMK allerdings direkt mit PISA in Verbindung gebracht wurden und dabei helfen sollten, die in PISA festgestellten Defizite zu beheben.

Zu beachten ist bei den Handlungsfeldern auch der politische Hintergrund, denn sie stellen nur Aspekte dar, auf die sich alle Kultusminister der Bundesländer einigen konnten. Strittige Maßnahmen, die beispielsweise nur von einer parteipolitischen Richtung vertreten wurden, wurden nicht aufgenommen, sodass beispielsweise Aussagen zum gegliederten Schulsystem komplett fehlen. Ein Vorteil dieser Verfahrensweise liegt darin, dass die Handlungsfelder Maßnahmen beschreiben, die in den Bundesländern bereits umgesetzt werden und Ergebnisse dadurch schneller sichtbar werden können. Als Nachteil erweist sich dagegen, dass, bedingt durch die zeitgleiche Darstellung der PISA-Ergebnisse, die Handlungsfelder nicht durch die Ergebnisse selbst bedingt sind, sondern bereits bestehende Entwicklungen und Maßnahmen aufnehmen (Tillmann, 2008).

Bündelung der Aktivitäten

### 3.5.2.2 Neue Schwerpunkte der KMK

Am 02.06.2006 hat die KMK eine Gesamtstrategie zum Bildungsmonitoring beschlossen. Innerhalb dieser Strategie werden vier miteinander verbundene Bereiche beschrieben, mit deren Hilfe Bildungsprozesse in Deutschland beobachtet und weiterentwickelt werden sollen. Dazu gehören:

- regelmäßige Teilnahme an internationalen Schulleistungsuntersuchungen,
- zentrale Überprüfung der Bildungsstandards im Ländervergleich,
- Vergleichsarbeiten (vgl. Studienbrief-Teil Vergleichsarbeiten),
- gemeinsame Bildungsberichterstattung.

Das übergeordnete Ziel dieser Gesamtstrategie liegt in der Beschaffung von Informationen, die für die Steuerung des Bildungssystems benötigt werden (vgl. Kapitel 3.1.1). Zudem sollen die so gewonnenen Erkenntnisse mit Maßnahmen der Unterrichts- und Qualitätsentwicklung verknüpft werden, sodass die pädagogische Arbeit an jeder einzelnen Schule davon profitieren kann (KMK, 2006).

Die Gesamtstrategie soll aber nicht eine umfassende Konzeption zur Weiterentwicklung des Bildungswesens darstellen. Vielmehr soll sie in bereits bestehende Beschlüsse der KMK eingebettet werden. Dazu gehören die im Folgenden beschriebenen Handlungsschwerpunkte und Arbeitsbereiche, die bereits nach PISA 2003 festgelegt wurden (KMK, 2006):

- Frühzeitige Förderung von Migranten und sozial Benachteiligten in der Bundesrepublik Deutschland
- Bereitstellung von Fortbildungskonzeptionen und -materialien zur kompetenz- bzw. standardbasierten Unterrichtsentwicklung
- Konzepte und Materialien für Deutsch als Aufgabe aller Fächer
- Aus- und Fortbildung der Lehrkräfte im Hinblick auf Verbesserung der Diagnosefähigkeit, Umgang mit Heterogenität, individuelle Förderung
- Erarbeitung eines flexibel im Unterricht einzusetzenden Aufgabenpools für die Fächer Deutsch und Mathematik

politischer Hintergrund

Durch die Gesamtstrategie wird deutlich, dass das Bildungsmonitoring mit konkreten Maßnahmen verknüpft werden muss. Dazu gehört neben der Unterrichts- und Qualitätsentwicklung auch die Unterstützung der Schulen (KMK, 2006).

Auf Grund der Ergebnisse von PISA 2006 und IGLU 2006 wurden von der KMK gemeinsam mit dem Bundesministerium für Bildung und Forschung im März 2008 Empfehlungen veröffentlicht. Darin werden die sieben Handlungsfelder aus dem Jahr 2001 als Grundlage für die Weiterentwicklung des Bildungssystems bestätigt und neue Schwerpunkte gesetzt (PISA-Konsortium Deutschland, 2008):

1. Stärkere Konzentration auf die Förderung leistungsschwacher Schülerinnen und Schüler insbesondere in der Sekundarstufe I
2. Durchlässigkeit des Bildungssystems verbessern, Übergänge erleichtern, Abschlüsse sichern
3. Den Unterricht weiter entwickeln, die Lehrkräfte qualifizieren

Die konkreten Veränderungen durch die Ergebnisse der Systemmonitoring-Studien und die damit verbundenen Entscheidungen zeigt das folgende Kapitel.

### 3.5.3 Auswirkungen des Systemmonitorings auf die Schulpraxis

Die Ergebnisse der Systemmonitoring-Studien und die sich daran anschließenden Diskussionen haben einen Einfluss auf „bildungspolitische Verläufe und schulische Innovationen“ (Tillmann, 2009) haben. Es wird durch die Studien möglich, dass bislang strittige Maßnahmen umgesetzt oder neue Pläne entwickelt werden. Dabei ist es auch möglich, dass die Ergebnisse als Legitimation für alternative Vorschläge verwendet werden, die bisher keine Mehrheit gefunden haben (Tillmann, 2009).

Als ein Beispiel für Auswirkungen des Systemmonitorings auf die Schulpraxis kann die Einführung zentraler Abiturprüfungen genannt werden. Dabei ist dies ein besonderes Beispiel, da die Einführung mit den Ergebnissen aus PISA begründet wurde, PISA selbst dazu allerdings gar keine Aussagen macht (Moschner/Kiper/Kattmann, 2003). Nähere Informationen dazu gibt Kapitel 3.5.3.4.

Eine Wirkung haben Studien im Rahmen des Systemmonitoring auf jeden Fall: Sie führen zu einer verstärkten Medienberichterstattung und erhöhen somit die Wichtigkeit der Bildungspolitik und den öffentlichen Druck auf die Bildungspolitiker.

Die folgenden Kapitel sollen in aller Kürze sechs Bereiche beschreiben, die sich durch die Systemmonitoring-Studien entwickelt beziehungsweise weiterentwickelt haben.

#### 3.5.3.1 Standardentwicklung

Unter ‚Bildungsstandards‘ wird international verstanden, normative Vorgaben für die Steuerung von Bildungssystemen zu erstellen. Dabei ergeben sich unterschiedliche Arten von Standards, einige von ihnen wurden bereits in Kapitel 3.2.3.4 beschrieben.

Bereits die Handlungsfelder der KMK aus dem Jahr 2001 beinhalteten die Entwicklung von Standards. Bis zu diesem Zeitpunkt gab es in Deutschland „ein Mischsystem, in dem mit der Vorgabe von Lehrplänen, mit Kompetenzprüfungen für das Personal und mit teilstandardisierten Abgangskontrollen, z. B. im Abitur oder in der Regulierung der Anforderungen in den mittleren Abschlüssen, die Qualität der Schularbeit gesichert werden soll(te)“ (Klieme u.a. 2007). Da durch PISA und TIMSS deutlich wurde, dass dieses Mischsystem nicht zum erwünschten Erfolg führte, wurde mit der Standardentwicklung begonnen.

Durch die Standards liegen ein bundesweites Referenzsystem und Bezugsgrößen für die Qualitätsentwicklung an deutschen Schulen vor (KMK, 2006). Auch wird es durch die Bildungsstandards möglich, eine Transparenz der schulischen Anforderungen herzustellen, einen kompetenzorientierten Unterricht zu entwickeln, Schülerinnen und Schüler gezielt zu fördern und erreichte Ergebnisse zu verdeutlichen (KMK, 2006). Somit gehört zur Standardentwicklung nicht nur die Erstellung der Standards, sondern auch die allgemeine Schulentwicklung und die interne und externe Evaluation (KMK, 2005). Durch die Standards wird zudem es möglich, schulische Abschlüsse unterschiedlicher Schularten zu vergleichen (KMK, 2005).

Die Bildungsstandards sollen in Zukunft zeitgleich mit den Systemmonitoring-Studien überprüft werden (vgl. Kapitel 3.3.6). Dazu werden vom IQB Testverfahren entwickelt, die die Standards operationalisieren und so eine Diagnose über die erreichten Kompetenzen ermöglichen. Die Ergebnisse der Standard-Überprüfung können dann für das weitere Bildungsmonitoring verwendet werden. Zudem können so ein kriterienorientierter Vergleich durchgeführt und Aussagen über die erworbenen Kompetenzen getroffen werden. Allerdings müssen dazu hohe Qualitätsstandards, wie beispielsweise bei PISA, eingehalten werden (Klieme u.a., 2007).

Die Charakteristika von Bildungsstandards lauten (KMK, 2005):

- die Grundprinzipien des jeweiligen Unterrichtsfaches werden aufgegriffen
- die fachbezogenen Kompetenzen werden beschrieben, einschließlich zugrunde liegender Wissensbestände, die Schülerinnen und Schüler bis zu einem bestimmten Zeitpunkt ihres Bildungsganges erreicht haben sollen
- es wird auf systematisches und vernetztes Lernen abgezielt, sie folgen dem Prinzip des kumulativen Kompetenzerwerbs
- sie beschreiben erwartete Leistungen im Rahmen von Anforderungsbereichen
- sie beziehen sich auf den Kernbereich des jeweiligen Faches (Basisqualifikationen) und geben den Schulen Gestaltungsräume für ihre pädagogische Arbeit
- sie weisen ein mittleres Anforderungsniveau auf (Regelstandards)
- Aufgabenbeispiele veranschaulichen die Standards

Überprüfung  
der Bildungs-  
standards

Hinzu kommt, dass die Bildungsstandards als ‚can do statements‘ formuliert werden. Sie beschreiben, was ein Schüler ganz genau können soll, um den Standard zu erreichen.

Can do state-  
ments

Von Bedeutung ist dabei, dass die Standards nicht die schulischen Lehr- und Lernprozesse beschreiben oder standardisieren. Stattdessen beschreiben sie die normative Erwartung, die innerhalb der Schule erfüllt werden soll. Auf welchen Wegen dies geschieht, wie die Lernzeit eingeteilt wird und wie mit personellen Ressourcen umgegangen wird, bleibt den einzelnen Bundesländern beziehungsweise Schulen überlassen. Das bedeutet, dass nur der Output der schulischen Bildung betrachtet wird. Damit führt die Standardentwicklung auch zu einer erhöhten Eigenverantwortung der einzelnen Schulen (KMK, 2005).

Bildungsstan-  
dards = Out-  
put- Orientie-  
rung

### 3.5.3.2 Einführung von Kernlehrplänen

Aus der Standardentwicklung ergibt sich die Konsequenz, dass auch die Lehrpläne angepasst werden müssen. Dabei hängt die genaue Ausgestaltung der Lehrpläne davon ab, wie stark die Output-Orientierung durchgeführt wird. Wird sie radikal durchgeführt, so verlieren Lehrpläne ihre Bedeutung als „strukturierendes Element von Unterricht“ (Klieme u.a., 2007). Lehrpläne, wie sie bis zum Zeitpunkt der Standardentwicklung ausgesehen haben, werden nicht mehr benötigt, wenn Standards die anzustrebenden Kompetenzen vorgeben. Bei radikaler Umsetzung kann sich jede Schule ihr eigenes Curriculum geben. In Deutschland wurde dieser radikale Weg nicht gewählt, stattdessen wurden Kernlehrpläne entwickelt. Sie beschreiben die Standards und Erwartungen, aber auch Themen und Inhalte, gelegentlich sogar Lernformen. In Kernlehrplänen werden die Bildungsstandards mit ihrer Leitfunktion an die bisherige Orientierungsfunktion von Lehrplänen gekoppelt. Zudem soll die Autonomie jeder einzelnen Schule gestärkt werden. Ein solcher Kernlehrplan entspricht dann den nationalen Standardvorgaben, kann aber auch in zeitliche Sequenzen eingeteilt werden und direkte Empfehlungen enthalten. So schließen sich Lehrpläne und Bildungsstandards nicht gegenseitig aus, vielmehr ergänzen sie sich. Sowohl die Standards als auch die Kernlehrpläne helfen dabei, die Qualität des deutschen Bildungswesens zu steigern und eine bessere Steuerung zu ermöglichen. Während die Standards dabei nur am Output orientiert sind, betrachten die Kernlehrpläne auch den Input. Dabei liegt es im Rahmen der Möglichkeiten, dass die Festlegung von Inhalten und Lernformen immer weiter von der Landes- an die Schulebene übergeben wird (Klieme u.a., 2007).

Inhalt der  
Kernlehr-  
pläne

Die Charakteristika von Kernlehrplänen lauten (Böttcher/Kalb, 2002 und Klieme u.a., 2007):

- Sie repräsentieren die Struktur von Bildung.
- Sie basieren auf Wissen (und wissensbasierten Kompetenzen).
- Sie bestimmen ein obligatorisches Fächergefüge.
- Sie nennen zentrale Themen und Inhalte, dabei aber nur das Minimum der zu Bearbeitenden.
- Sie sind offen für fachinterne Vertiefung und die thematische Koppelung von Lehrgegenständen.
- Sie bezeichnen erwartete Kompetenzen der Adressaten schulischer Arbeit. Dabei werden nur Ziele formuliert, die für alle Schülerinnen und Schüler erreichbar sind.
- Sie sind klar, eindeutig und verbindlich formuliert.
- Sie erlauben die Profilbildung von Einzelschulen.
- Sie füllen nicht mehr als 60 Prozent der Lernzeit an einer durchschnittlichen Schule.
- Sie müssen regelmäßig evaluiert und verändert werden.

Ein Problem der Kernlehrpläne liegt darin, dass das Bildungssystem auf ein anderes System umgestellt werden muss. Dadurch werden vor allem die Akteure im Bildungswesen vor Herausforderungen gestellt. Zunächst müssen sich daher alle an die neuen Lehrpläne gewöhnen und den Umgang mit ihnen erproben, zudem muss bei allen Akteuren eine große Akzeptanz für die Kernlehrpläne hergestellt werden.

Problem der  
Kernlehrpläne



### 3.5.3.3 Stärkung der Sprachkompetenz

Sprachliche Kompetenzen sind unbestritten wichtig für die Entwicklung und den beruflichen Erfolg eines Menschen. Bereits die Handlungsfelder der KMK aus dem Jahr 2001 beinhalteten daher die Förderung der Sprachkompetenz. Durch die Studien des Systemmonitorings, allen voran PISA, wurde deutlich, dass das deutsche Schulsystem nicht in der Lage ist, allen Schülerinnen und Schülern eine angemessene sprachliche Kompetenz in der Unterrichtssprache Deutsch zu vermitteln. Daher wurden verschiedene Programme aufgelegt, die die Stärkung der Sprachkompetenz zum Inhalt hatten. Dabei werden Schülerinnen und Schüler mit Migrationshintergrund, aber auch Schülerinnen und Schüler mit Deutsch als Erstsprache gefördert (Gogolin, 2008). In diesem Unterkapitel sollen beispielhaft drei Projekte zur Förderung der Sprachkompetenz dargestellt werden: FörMig, ProLesen und Delfin4, dabei wird davon ausgegangen, dass in allen Bundesländern weitere beziehungsweise ähnliche Projekte durchgeführt werden.

#### FörMig

Das Projekt „Förderung von Kindern und Jugendlichen mit Migrationshintergrund“ (FörMig) wurde von der Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (BLK)<sup>20</sup> als große Modellinitiative als Reaktion auf das schlechte Abschneiden aller Schülerinnen und Schüler im Bereich der Lesekompetenz in PISA 2000 gestartet. Das Projekt begann Ende 2004 und endete Ende 2009. Ziel des Projektes war es, den beteiligten Bundesländern zu ermöglichen, ihre innovativen Ansätze zur Sprachförderung zu optimieren und somit für einen „Transfer guter Praxis zu sorgen“. Damit identifiziert werden kann, was genau als ‚gute Praxis‘ gelten kann, wurden Evaluationen der einzelnen Projekte durchgeführt. Zudem wurde eine zentrale Evaluation erstellt. Sie verbindet die sprachlichen Entwicklungen der beteiligten Kinder mit ihren individuellen Voraussetzungen, ihren Kontextbedingungen und den Handlungsbedingungen der beteiligten Institutionen. Die Ergebnisse der zentralen Evaluation sollen einerseits für die Bildungsplanung verwendet werden, andererseits sollen sie in die praktischen Handlungen übernommen werden. Dabei betrachtet FörMig vor allem die Übergänge von einem Bildungsweg in den anderen. Einerseits wird genau dieser Übergang untersucht und Möglichkeiten zur Gestaltung, wie beispielsweise Lernbiografien oder Sprachlerntagebücher, werden entwickelt, andererseits beschäftigt sich das Projekt mit der Sprachdiagnostik. Dabei ist das Projekt nicht nur auf die deutsche Sprache beschränkt, vielmehr wird davon ausgegangen, dass alle Sprachen eines Kindes wichtig für dessen weitere Entwicklung sind. Daher beinhaltet FörMig auch Einheiten, in denen die Kompetenzen in der Erstsprache der Kinder und Jugendlichen gefördert werden. Die einzelnen Projekte werden immer so konzipiert, dass Partnerschaften gebildet werden. Diese können mit Eltern, Vereinen, Bibliotheken oder auch Betrieben zustande kommen. Das abschließende Ziel des Gesamtprojektes ist es dabei nicht nur, Förderkonzepte für Deutsch zu entwickeln und zu erproben, sondern auch dazu beizutragen, dass diese Konzepte in das deutsche Bildungssystem integriert werden (Gogolin, 2008).

#### ProLesen

Das Projekt „ProLesen. Auf dem Weg zur Leseschule“ ist ein Projekt der KMK und basiert ebenfalls auf dem ersten KMK-Handlungsfeld aus dem Jahr 2001. Inhalt des Projektes ist die Förderung von Deutsch als Unterrichtssprache als Aufgabe aller Fächer. Dazu sollen in zehn verschiedenen Modulen für alle Schularten der Primar- und Sekundarstufen Konzepte und Materialien gesichtet, gesammelt, überarbeitet und neu entwickelt werden. Abschließend sollen diese dann aufeinander abgestimmt und zu einem Gesamtkonzept schulischer Leseförderung zusammengefasst werden. Besonders Förderkonzepte für die in PISA erkannten Risikogruppen (Lesekompetenz: Jungen und Jugendliche mit Migrationshintergrund) sollen erstellt werden. Zusätzlich soll die Diagnose- und Förderkompetenz der Lehrkräfte verbessert werden. Alle Ergebnisse des Projektes sollen gemeinsam veröffentlicht und den Schulen zugänglich gemacht werden. Abschließendes Ziel des Projekts ist es, Problembereiche zu beschreiben und Beispiele guter Praxis und Materialien zur Verfügung zu stellen. Dabei sollen die beschriebenen Maßnahmen konkret ihre Wirkungsmöglichkeiten darstellen und leicht benutzbar sein. Zusätzlich werden Beispielaufgaben angefertigt, die gleichzeitig auch der Lernstandserhebung dienen und somit den Lehrkräften bei der Diagnose behilflich sein können (Staatsinstitut für Schulqualität und Bildungsforschung München, 2008).

#### Delfin 4

In Nordrhein-Westfalen wird mit dem Projekt „Delfin4“ zwei Jahre vor der Einschulung die Sprachkompetenz der Kinder überprüft. Das Projekt beruht darauf, dass im Jahr 2008 die kontinuierliche sprachliche Entwicklung aller Kinder des Landes gesetzlich festgehalten wurde. Alle Kinder, auch diejenigen, die keine Kindertageseinrichtung besuchen, werden in einem zweistufigen Verfahren getestet. Dabei ist wichtig, dass die Eltern umfassend informiert werden und die Ergebnisse zeitnah erhalten. Für dieses Verfahren arbeiten die Kindertageseinrichtungen mit den nahegelegenen Grundschulen zusammen. Alle Kinder, für die ein Sprachförderbedarf festgestellt wurde, erhalten Sprachförderung. Die Teilnahme an der Sprachförderung ist verpflichtend. Das Land Nordrhein-Westfalen stellt pro an der Sprachförderung teilnehmendem Kind Geld zur Verfügung. Die Sprachförderung wird innerhalb der Kindertagesstätte durchgeführt und soll in die grundständige Sprachförderung der Einrichtung

<sup>20</sup> Auf Grund der Föderalismusreform existiert die BLK nicht mehr. Allerdings haben sich die zehn am Projekt teilnehmenden Bundesländer darauf geeinigt, das Projekt in eigener Verantwortung zu Ende zu führen.

eingebettet werden. Die Ausführung und Gestaltung der Sprachförderung liegt in der Hand des Einrichtungsträgers, durchgeführt wird sie von geeigneten pädagogischen Fachkräften (Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2008).

### 3.5.3.4 Ganztagsschulen

Der Ausbau schulischer und außerschulischer Ganztagsangebote ist bereits Bestandteil der KMK-Handlungsfelder aus dem Jahr 2001. Die Aufnahme in die Handlungsfelder hat in Kombination mit den ersten PISA-Ergebnissen in mehreren Bundesländern dazu geführt, dass das Ganztags-Schulsystem intensiv ausgebaut wurde. Bis zu diesem Zeitpunkt waren die meisten deutschen Schulen Halbtagschulen und endeten gegen 13 Uhr. In anderen Staaten wie beispielsweise England und Frankreich ist der Ganztagschulbetrieb dagegen schon lange selbstverständlich (Tillmann, 2008).

Ganztagschulen

Dabei bietet PISA keine wissenschaftlichen Ergebnisse darüber, dass die Einführung von Ganztagschulen die Probleme des deutschen Bildungssystems lösen kann. Trotzdem waren Politiker aller Parteien der Überzeugung, dass die Einführung von Ganztagschulen viele Probleme lösen kann. Erleichtert wurde die Einführung dadurch, dass die Bundesregierung Geld für zusätzliche 10.000 Ganztagschulen im ganzen Land zur Verfügung stellte und so der Anreiz für die Bundesländer größer wurde, auch wenn sie selbst noch investieren mussten (Tillmann, 2009).

Im Ganztagsbereich wird durch eine Zusammenarbeit von Schule, Jugendhilfe, Kultur und Sport ein abwechslungsreiches Programm angeboten. Neben einer Hausaufgabenbetreuung werden beispielsweise Aktivitäten im Bereich Sport und Spiel, Kreativität, Sprachförderung für Kinder mit Migrationshintergrund, Entspannung und Konzentration angeboten. Zudem erhalten die Kinder ein Mittagessen. Die Teilnahme am Ganztag ermöglicht den Kindern, sich außerhalb des Unterrichts auszuprobieren und zu erkennen, in welchen Bereichen ihre Fähigkeiten liegen. Die Schülerinnen und Schüler können somit durch ein Ganztagsangebot besser individuell gefördert werden (Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2007).

### 3.5.3.5 SINUS / SINUS transfer

Das Projekt „Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts“ (SINUS) wurde von der BLK auf Grund der Ergebnisse von TIMSS (vgl. Kapitel 3.4.2.1) aufgelegt. Zur Vorbereitung des Programms wurde 1997 ein Gutachten erstellt, das die Problemzonen des naturwissenschaftlichen Unterrichts beschrieb. Auf Grundlage dieses Gutachtens wurde dann die Konzeption von SINUS erstellt.

SINUS

Die Konzeption des Projekts wird (nach BLK, 2001) wie folgt beschrieben:

„Prozesse der Qualitätssicherung und Optimierung von Lehren und Lernen in den mathematisch-naturwissenschaftlichen Fächern sollen auf der Ebene der Schule in Gang gesetzt und mit dem Ziel gestützt werden, diesen eine eigene Dynamik zu geben, die über den Modellversuch hinaus trägt.“

Das Konzept beruht auf der Annahme, dass Veränderungen im Bereich Schule sich nur entwickeln und bestehen können, wenn sie von den Lehrkräften angenommen werden und diese sie in ihre Handlungsrouinen aufnehmen. Dabei setzt das Programm an den Stärken des mathematisch-naturwissenschaftlichen Themenfeldes an und schlägt Module vor, die in den Schulen und gebildeten Schulnetzen ausgewählt und bearbeitet werden können. Alle Module betreffen jeweils einen konkreten Problembereich des mathematisch-naturwissenschaftlichen Unterrichts und enthalten Hinweise auf die Möglichkeiten der Bearbeitung. Das Grundprinzip dabei ist, dass Lehrkräfte zusammenarbeiten sollen, zunächst innerhalb der Fachgruppe einer Schule, längerfristig auch über die eigene Schule hinaus.

Die einzelnen Module müssen von jeder Schule entsprechend der lokalen und regionalen Bedingungen interpretiert und konkretisiert werden.

Vor Beginn des Projekts wurden zu allen elf Modulen Handreichungen erstellt. Darin werden die Problemzonen genauer beschrieben und der Forschungsstand skizziert. Zudem werden Möglichkeiten der Bearbeitung und eventuelle Schwierigkeiten beschrieben. Beispiele sind ebenfalls enthalten. Aufgabe der Schule ist es, die Module in der praktischen Arbeit umzusetzen, dabei werden sie von wissenschaftlicher Seite unterstützt. Ihre Arbeit soll protokolliert werden, dafür stellt das Projekt Vorlagen zur Verfügung. Zudem wurde eine Internetplattform eingerichtet, auf der Informationen ausgetauscht und Kooperationen gepflegt werden können. Es ist ebenfalls Aufgabe der Schule, mit bereitgestellten Instrumenten den veränderten Unterricht zu evaluieren. Neben dieser internen Evaluation wurde das Projekt zusätzlich extern evaluiert (BLK, 2001).



Nach Ablauf des Projektzeitraums (April 1998 bis März 2001) wurde das Projekt unter dem Namen „SINUS-Transfer“ weitergeführt.

### 3.5.3.6 For.mat

For.mat

Das Projekt „Bereitstellung von Fortbildungskonzeptionen und -materialien zur kompetenz- bzw. standardbasier-ten Unterrichtsentwicklung, vor allem Lesen, Geometrie, Stochastik“ ist ein Projekt der KMK, das seit Januar 2007 unter der Federführung des Landes Rheinland-Pfalz durchgeführt wurde. Innerhalb des Projekts entwickelten Personen aus allen 16 Bundesländern Materialien und Konzeptionen und stellten diese für die allgemeine Nutzung bereit. Dabei sollte ein sinnhafter Anschluss an das Projekt SINUS (vgl. Kapitel 3.5.3.5) erreicht werden. Der Unterschied zu SINUS besteht darin, dass alle Fächer betrachtet werden, für die durch die KMK Bildungsstandards beschlossen wurden. Ziel des Projekts war „die systematische Entwicklung und Qualifizierung von Fachkonferenzen und Fachgruppen zu professionellen schulinternen Lerngemeinschaften in den Bildungsstandardfächern Deutsch, erste Fremdsprache, Mathematik und Naturwissenschaften“

Das Projekt teilt sich in zwei Teilprojekte auf:

- Teilprojekt 1: Einrichtung von vier Arbeitsgruppen, diese entwickeln „fachbezogene Materialien und Konzeptionen für die kompetenzorientierte Unterrichtsentwicklung in professionellen Fachgruppen“ und stellen diese zur Verfügung
- Teilprojekt 2 hat einen „Fokus auf Kompetenzprofile und Qualifizierungskonzepte für Beraterinnen und Berater zur Unterstützung einer kompetenz- und standardbasierten Unterrichtsentwicklung“ ([www.kmk-format.de](http://www.kmk-format.de), Stand 12.2.2010)

### 3.5.3.7 UDiKom

UDiKom

Das Projekt UDiKom, in dessen Rahmen der vorliegende Studienbrief erstellt wurde, greift das sechste Handlungsfeld der KMK von 2001 auf: „Stärkung der diagnostischen und methodischen Kompetenzen der Lehrkräfte“. Durch die Studienbrief-Module und die zusätzlichen Materialien soll vermittelt werden, wie diagnostische Instrumente für die eigene Arbeit verwendet werden können.

### 3.5.3.8 Bildungsberichterstattung

Bildungsbericht-  
erstattung

Die regelmäßige Bildungsberichterstattung ist Teil der KMK-Gesamtstrategie zum Bildungsmonitoring. Dabei sollen keine neuen empirischen Studien durchgeführt werden, sondern vorliegende Daten zu allen Bildungsbereichen unter der Leitidee „Bildung im Lebenslauf“ im Zusammenhang dargestellt und als Grundlage für politische Entscheidungen und für gesellschaftliche Diskussion zur Verfügung. Diese Anforderung impliziert neben der zur Verfügungstellung der Daten die Analyse, Interpretation und gesellschaftliche Einordnung. Durchgeführt wird diese Arbeit von den folgenden Institutionen: Deutsches Institut für Internationale Pädagogische Forschung (DIPF) in Kooperation mit dem Deutschen Jugendinstitut (DJI), dem Hochschul-Informationssystem GmbH (HIS), dem Soziologischen Forschungsinstitut an der Universität Göttingen (SOFI) sowie den Statistischen Ämtern des Bundes und der Länder (StBA, StLÄ).

Das Ziel des Berichts liegt darin, sowohl den Zustand des deutschen Bildungssystems, die Entwicklung über die letzten Jahre als auch die aktuellen Herausforderungen zu dokumentieren. Dabei sollen die folgenden Aspekte in konzentrierter Form betrachtet werden:

- Aktuelle Situation des deutschen Bildungssystems
- Leistungsfähigkeit des deutschen Bildungssystems
- Wichtigste Problemlagen des deutschen Bildungssystems
- Bildungsprozesse im Lebenslauf
- Entwicklung des deutschen Bildungswesens im internationalen Vergleich

Der Vorteil gegenüber vielen einzelnen Berichten besteht darin, „dass die verschiedenen Bildungsbereiche in ihrem Zusammenhang gesehen, analysiert und dargestellt werden. Auf diesem Weg lassen sich übergreifende Probleme im deutschen Bildungswesen für Bildungspolitik und Öffentlichkeit sichtbar machen sowie handlungs- und steuerungsrelevante Informationen für Politik und Verwaltung gewinnen“. Der Anspruch des nationalen Bildungsberichts ist somit hoch, er will eine „Gesamtschau des Bildungssystems“ geben und greift dabei auf bereits bestehende Daten zurück, die durch „eine überschaubare Zahl von Indikatoren verdichtet“ werden. ([http://www.bildungsbericht.de/daten2008/bb\\_2008.pdf](http://www.bildungsbericht.de/daten2008/bb_2008.pdf), Stand 10.02.10)

Bisher wurden in den Jahren 2006 und 2008 Bildungsberichte erstellt. Die zentralen Themen, die in jedem Jahr betrachtet werden, sind:

- Bildung im Spannungsfeld veränderter Rahmenbedingungen
- Grundinformationen zu Bildung in Deutschland
- Frühkindliche Bildung, Betreuung und Erziehung
- Allgemeinbildende Schule und non-formale Lernwelten im Schulalter
- Berufliche Ausbildung
- Hochschule
- Weiterbildung und Lernen im Erwachsenenalter
- Wirkung und Erträge von Bildung

Zudem wurden in jedem Bildungsbericht Schwerpunktanalysen angestellt. Im Jahr 2006 war dies der Bereich ‚Migration‘, im Jahr 2008 ‚Übergänge im Anschluss an den Sekundarbereich I‘.

Sehr ausführliche Informationen stehen unter [www.bildungsbericht.de](http://www.bildungsbericht.de) zur Verfügung, hier können auch die einzelnen Berichte als PDF-Dokumente heruntergeladen werden.

### 3.5.4 Weiterführende Literatur

Kohler (2005) beschreibt wissenschaftliche Untersuchungen zur Rezeption internationaler Schulleistungstudien. Aktuelle Informationen über die verschiedenen vorgestellten Projekte können dem Internet entnommen werden.

*FörMig:* <http://www.blk-foermig.uni-hamburg.de/>

*ProLesen:* <http://www.leseforum.bayern.de/index.asp?MNav=6>

*Delfin 4:* <http://www.delfin4.fb12.uni-dortmund.de/>

*SINUS/SINUS-Transfer:* <http://blk.mat.uni-bayreuth.de/indexblk.html>

<http://www.sinus-transfer.de/>

### 3.5.5 Verständnis-Aufgaben und Diskussionspunkte

1. Leiten Sie eine eigene Projektidee aus den vorgestellten Handlungsfeldern/ Schwerpunkten ab. Überprüfen Sie dabei, ob sich Ihre Idee mit einer der vorgestellten Untersuchungen legitimieren lässt.
2. Beschreiben Sie, analog zum Beispiel des Gymnasiallehrers, welche Fragen sich Eltern und Bildungspolitiker stellen könnten.
3. Überlegen Sie, in welchen Bereichen innerhalb der sieben Handlungsfelder der KMK von 2001 Veränderungsbedarf an Ihrer Bildungseinrichtung bestehen könnte.
4. Suchen Sie (z.B. im Internet) nach weiteren Projekten, die in den Handlungsfeldern und Schwerpunkten der KMK impliziert werden.
5. Beschreiben Sie, wie die acht beschriebenen Veränderungsbereiche in den Kapiteln 3.5.3.1 bis 3.5.3.8 die Ergebnisse von zukünftigen Untersuchungen verbessern können.

### 3.6 Literaturverzeichnis

Ackeren, Isabell van (2006); *Internationale Vergleichsuntersuchungen*. Studienbrief für die Technische Universität Kaiserslautern; Fernstudium Schulmanagement, Kaiserslautern.

Albert, Ruth / Koster, Cor J. (2002); *Empirie in Linguistik und Sprachlehrforschung*; Gunter Narr Verlag, Tübingen

Arnold, Karl-Heinz (2001); Qualitätskriterien für die Messung von Schulleistungen; in: Weinert, Franz E. (Hg.); *Leistungsmessungen in Schulen*; Beltz Verlag, Weinheim u.a., S. 117-130

Autorengruppe Bildungsberichterstattung (2008); *Bildung in Deutschland 2008*. Ein indikatorengestützter Bericht mit einer Analyse zu Übergängen im Anschluss an den Sekundarbereich I; Bertelsmann; Bielefeld.  
[http://www.bildungsbericht.de/daten2008/bb\\_2008.pdf](http://www.bildungsbericht.de/daten2008/bb_2008.pdf) (Stand 10.02.2010)

Baumert, Jürgen / Lehmann, Rainer (1997); TIMSS - mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich - deskriptive Befunde; Leske + Budrich Verlag, Opladen

Baumert, Jürgen / Bos, Wilfried / Watermann, Rainer (1998); TIMSS/III: Schülerleistungen in Mathematik und den Naturwissenschaften am Ende der Sekundarstufe II im internationalen Vergleich ; Zusammenfassung deskriptiver Ergebnisse; Max-Planck-Institut für Bildungsforschung, Berlin

Baumert, Jürgen / Bos, Wilfried / Lehmann, Rainer (2000); TIMSS/III: Mathematische und naturwissenschaftliche Grundbildung am Ende der Schullaufbahn, Band 1; Leske + Budrich Verlag, Opladen

Beck, Bärbel / Klieme, Eckhard (Hg.) (2007); *Sprachliche Kompetenzen - Konzepte und Messung ; DESI-Studie (Deutsch Englisch Schülerleistungen International)*; Beltz Verlag, Weinheim u.a.

BLK - Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (2001); *Die Grundkonzeption des BLK-Programms „Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts“*;  
<http://blk.mat.uni-bayreuth.de/programm/konzeption.html> (Zugriff am: 30.09.2009)

Böttcher, Wolfgang / Kalb, Peter E. (2002); *Kerncurriculum. Was Kinder in der Schule lernen sollen*. Eine Streitschrift; Beltz Verlag, Weinheim u.a.

Bos, Wilfried / Postlethwaite, Thomas Neville (2001); *Internationale Schulleistungsforschung: Ihre Entwicklungen und Folgen für die deutsche Bildungslandschaft*; in: Weinert, Franz E. (Hg.); *Leistungsmessungen in Schulen*; Beltz Verlag, Weinheim u.a., S. 251-267

Bos, Wilfried / Lankes, Eva-Maria / Prenzel, Manfred / Schwippert, Knut / Walther, Gerd / Valtin, Renate (Hg.) (2003); *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*; Waxmann Verlag, Münster

Bos, Wilfried / Hornberg, Sabine / Arnold, Karl-Heinz / Faust, Gabriele / Fried, Lilian / Lankes, Eva-Maria / Schwippert, Knut / Valtin, Renate (Hg.) (2007); *IGLU 2006 – Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*; Waxmann Verlag, Münster

Bos, Wilfried / Bensen, Martin / Baumert, Jürgen / Prenzel, Manfred / Selter, Christoph / Walther, Gerd (Hg.) (2008); *TIMSS 2007 - Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*; Waxmann Verlag, Münster

Bos, Wilfried / Hornberg, Sabine / Arnold, Karl-Heinz / Faust, Gabriele / Fried, Lilian / Lankes, Eva-Maria / Schwippert, Knut / Valtin, Renate (Hg.) (2008); *IGLU-E 2006 - Die Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich*; Waxmann Verlag, Münster

DESI-Konsortium (Hg.) (2008); *Unterricht und Kompetenzerwerb in Deutsch und Englisch – Ergebnisse der DESI-Studie*; Beltz Verlag, Weinheim u.a.

Deutsches PISA-Konsortium (Hg.) (2001); *PISA 2000 - Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*; Leske + Budrich Verlag, Opladen

Deutsches PISA-Konsortium (Hg.) (2002); *PISA 2000 - die Länder der Bundesrepublik Deutschland im Vergleich*; Leske + Budrich Verlag, Opladen

Deutsches PISA-Konsortium (2003); PISA 200: ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland; Leske + Budrich Verlag, Opladen

Diehl, Joerg M. / Kohr, Heinz U. (2004); Deskriptive Statistik; 13. Auflage; Verlag Dietmar Klotz, Eschborn

Döbert, Hans et al. (2009); Das Indikatoren-Konzept der nationalen Bildungsberichterstattung in Deutschland; in: Tippelt, Rudolf (Hg.); Steuerung durch Indikatoren - methodologische und theoretische Reflektionen zur deutschen und internationalen Bildungsberichterstattung; Verlag Barbara Budrich, Opladen, S. 207-272

Eichler, Wolfgang (2003); Die PISA-Nachfolgestudie DESI: „Deutsch-Englische Sprachkompetenz von Schülerinnen und Schülern der neunten Jahrgangsstufe – International“; in: Moschner, Barbara / Kiper, Hanna / Kattmann, Ulrich (Hg.); PISA 2000 als Herausforderung. Perspektiven für Lehren und Lernen; Schneider Verlag Hohengehren, Baltmannsweiler, S. 157-171

For.Mat; Bereitstellung von Fortbildungskonzeptionen und -materialien zur kompetenz- bzw. standardbasierten Unterrichtsentwicklung, vor allem Lesen, Geometrie, Stochastik. [www.kmk-format.de](http://www.kmk-format.de) (Stand 12.02.2010)

Gogolin, Ingrid (2008); „Förderung von Kindern und Jugendlichen mit Migrationshintergrund FörMig“ – ein länderübergreifendes Programm zur Optimierung der Sprachbildung; in: Gesellschaft. Wirtschaft. Politik (GWP); Heft 1/2008, Verlag Barbara Budrich, S. 65-75

Grotlüschen, Anke / Linde, Andrea (Hg.) (2007); Literalität, Grundbildung oder Lesekompetenz? Beiträge zu einer Theorie-Praxis-Diskussion; Waxmann Verlag, Münster

Heller, Kurt A. / Hany, Ernst A. (2001); Standardisierte Schulleistungsmessungen; in: Weinert, Franz E. (Hg.); Leistungsmessungen in Schulen; Beltz Verlag, Weinheim u.a., S. 87-101

Helmke, Andreas / Schrader, Friedrich-Wilhelm (1998). Determinanten der Schulleistung; in: Rost, Detlef H. (Hg.); Handwörterbuch Pädagogische Psychologie; Beltz Verlag, Weinheim u.a., S. 60-67

Kiper, Hanna (2003a); PISA-Ergänzungsstudie – eine erste Zusammenfassung ihrer Ergebnisse; in: Moschner, Barbara / Kiper, Hanna / Kattmann, Ulrich (Hg.); PISA 2000 als Herausforderung. Perspektiven für Lehren und Lernen; Schneider Verlag Hohengehren, Baltmannsweiler, S. 39-52

Kiper, Hanna (2003b); literacy versus Curriculum?; in: Moschner, Barbara / Kiper, Hanna / Kattmann, Ulrich (Hg.); PISA 2000 als Herausforderung. Perspektiven für Lehren und Lernen; Schneider Verlag Hohengehren, Baltmannsweiler, S. 65-86

Kiper, Hanna / Kattmann, Ulrich (2003); Basiskompetenzen im Vergleich – Übersicht über Ergebnisse der PISA-Studie 2000; in: Moschner, Barbara / Kiper, Hanna / Kattmann, Ulrich (Hg.); PISA 2000 als Herausforderung. Perspektiven für Lehren und Lernen; Schneider Verlag Hohengehren, Baltmannsweiler, S. 15-37

Klieme, Eckhard / Baumert, Jürgen / Köller, Olaf (2000); Mathematische und naturwissenschaftliche Grundbildung: Konzeptuelle Grundlagen und die Erfassung und Skalierung von Kompetenzen; in: Baumert, Jürgen / Bos, Wilfried / Lehmann, Rainer; TIMSS/III: Mathematische und naturwissenschaftliche Grundbildung am Ende der Schullaufbahn, Band 1; Leske + Budrich Verlag, Opladen, S. 85-133

Klieme, Eckhard / Baumert, Jürgen (2001); TIMSS-Impulse für Schule und Unterricht. Forschungsbefunde, Reforminitiativen, Praxisberichte und Video-Dokumente; Bundesministerium für Bildung und Forschung (Hg.); Bonn

Klieme, Eckhard / Neubrand, Michael / Lüdtke, Oliver (2001); Mathematische Grundbildung: Testkonzeption und Ergebnisse; in: Deutsches PISA-Konsortium (Hg.); PISA 2000 - Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich; Leske + Budrich Verlag, Opladen; S. 141-190

Klieme, Eckhard (2006); Zusammenfassung zentraler Ergebnisse der DESI-Studie; [http://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2006/2006\\_03\\_01-DESI-Ausgewaehlte-Ergebnisse.pdf](http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2006/2006_03_01-DESI-Ausgewaehlte-Ergebnisse.pdf) (Zugriff am: 30.09.2009)

Klieme, Eckhard / Avenarius, Hermann / Blum, Werner / Döbrich, Peter / Gruber, Hans / Prenzel, Manfred / Reiss, Kristina / Riquarts, Kurt / Rost, Jürgen / Tenorth, Heinz-Elmar / Vollmer, Helmut J. (2007); Zur Entwicklung nationaler Bildungsstandards, Bundesministerium für Bildung und Forschung (Hg.), Bonn, Berlin

KMK – Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2005); Bildungsstandards der Kultusministerkonferenz. Erläuterungen zur Konzeption und Entwicklung; Wolters Kluwer Verlag, München

KMK – Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2006); Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring; Wolters Kluwer Verlag, München

KMK – Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2009); Kompetenzstufenmodell zu den Bildungsstandards im Kompetenzbereich Sprechen und Zuhören – hier Zuhören – für den Mittleren Schulabschluss

Köller, Olaf / Baumert, Jürgen / Bos, Wilfried (2001); TIMSS - Third International Mathematics and Science Study: Dritte internationale Mathematik- und Naturwissenschaftsstudie; in: Weinert, Franz E. (Hg.); Leistungsmessungen in Schulen; Beltz Verlag, Weinheim u.a., S.269-284

Köller, Olaf (2008); Bildungsstandards in Deutschland: Implikationen für die Qualitätssicherung und Unterrichtsqualität; in: Zeitschrift für Erziehungswissenschaft; 10. Jahrgang, Sonderheft 9/2008, S. 47-59

Kohler, Britta (2005); Rezeption internationaler Schulleistungstudien; Waxmann Verlag, Münster

Kraus, Josef (2003); TIMSS, PISA, IGLU und Co. – Fakten und Legenden; in: Kraus, Josef / Schmoll, Heike / Gauger, Jörg-Dieter; Von Timss zu IGLU. Eine Nation wird vermessen; Sankt Augustin, S. 7-72

Maritzen, Norbert (2008); Bildungsmonitoring – Systemevaluation zur Sicherung von Qualitätsstandards; in: Landesinstitut für Schule und Medien Berlin-Brandenburg (Lisum) / Kunst und Kultur (bm:ukk, Österreich) / Bundesministerium für Unterricht (Schweiz) / Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK) (Hg.); Bildungsmonitoring, Vergleichsstudien und Innovationen: Von evidenzbasierter Steuerung zur Praxis; Berliner Wissenschafts-Verlag, Berlin, S. 109-124

Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (2007); GanzTag, Ausgabe 01/2007, Düsseldorf

Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (2008); Feststellung des Sprachstands zwei Jahre vor der Einschulung. Fachinformation zum verfahren 2009, Düsseldorf

Moschner, Barbara / Kiper, Hanna / Kattmann, Ulrich (Hg.) (2003); PISA 2000 als Herausforderung. Perspektiven für Lehren und Lernen; Schneider Verlag Hohengehren, Baltmannsweiler

OECD (2006); Assessing Scientific, Reading and Mathematical literacy – A Framework for PISA 2006; OECD-Publishing, Paris

PISA-Konsortium Deutschland (Hg.) (2004); PISA 2003 – Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs; Waxmann Verlag, Münster

PISA-Konsortium Deutschland (Hg.) (2005); PISA 2003 - Der zweite Vergleich der Länder in Deutschland - was wissen und können Jugendliche?; Waxmann Verlag, Münster

PISA-Konsortium Deutschland (Hg.) (2006); PISA 2003 - Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres; Waxmann Verlag, Münster

PISA-Konsortium Deutschland (Hg.) (2007); PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie; Waxmann Verlag, Münster

PISA-Konsortium Deutschland (Hg.) (2008); PISA 2006 in Deutschland. Die Kompetenzen der Jugendlichen im dritten Ländervergleich; Waxmann Verlag, Münster

PISA-Konsortium Deutschland (Hg.) (2010); PISA 2009. Bilanz nach einem Jahrzehnt; Waxmann Verlag, Münster

Rauschenbach, Thomas (2009); Informelles Lernen. Möglichkeiten und Grenzen der Indikatorisierung; Tippelt, Rudolf (Hg.); Steuerung durch Indikatoren - methodologische und theoretische Reflektionen zur deutschen und internationalen Bildungsberichterstattung; Verlag Barbara Budrich, Opladen, S. 35-53

Rolff, Hans-Günter (2008); Konsequenzen aus Schulleistungsstudien und ihre Umsetzung auf Schulebene; in: Landesinstitut für Schule und Medien Berlin-Brandenburg (Lisum) / Kunst und Kultur (bm:ukk, Österreich) / Bundesministerium für Unterricht (Schweiz) / Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK) (Hg.); Bildungsmonitoring, Vergleichsstudien und Innovationen: Von evidenzbasierter Steuerung zur Praxis; Berliner Wissenschafts-Verlag, Berlin, S. 147-159

Specht, Werner (2008); Innovation durch Evaluation?; in: Landesinstitut für Schule und Medien Berlin-Brandenburg (Lisum) / Kunst und Kultur (bm:ukk, Österreich) / Bundesministerium für Unterricht (Schweiz) / Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK) (Hg.); Bildungsmonitoring, Vergleichsstudien und Innovationen: Von evidenzbasierter Steuerung zur Praxis; Berliner Wissenschafts-Verlag, Berlin, S. 41-52

Staatsinstitut für Schulqualität und Bildungsforschung München (2008); Projektplan ProLesen. Auf dem Weg zur Leseschule – Konzepte und Materialien zur Leseförderung als Aufgabe aller Fächer; München

Stadelmann, Willi (2008); Konsequenzen aus Schulleistungsstudien; in: Landesinstitut für Schule und Medien Berlin-Brandenburg (Lisum) / Kunst und Kultur (bm:ukk, Österreich) / Bundesministerium für Unterricht (Schweiz) / Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK) (Hg.); Bildungsmonitoring, Vergleichsstudien und Innovationen: Von evidenzbasierter Steuerung zur Praxis; Berliner Wissenschafts-Verlag, Berlin, S. 169-175

Stanat, Petra (2008); Entstehung und Umsetzung von Innovationen im Bildungssystem als Konsequenz aus Bildungsmonitoring, Bildungsberichterstattung und vergleichenden Schulleistungsstudien – Möglichkeiten und Grenzen; in: Landesinstitut für Schule und Medien Berlin-Brandenburg (Lisum) / Kunst und Kultur (bm:ukk, Österreich) / Bundesministerium für Unterricht (Schweiz) / Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK) (Hg.); Bildungsmonitoring, Vergleichsstudien und Innovationen: Von evidenzbasierter Steuerung zur Praxis; Berliner Wissenschafts-Verlag, Berlin, S. 11-23

Tillmann, Klaus-Jürgen (2008) (Hg.); PISA als bildungspolitisches Ereignis  
Fallstudien in vier Bundesländern; Verlag für Sozialwissenschaften, Wiesbaden

Tillmann, Klaus-Jürgen (2009); Was leistet die PISA-Studie zur Steuerung des Bildungssystems?; in: Tippelt, Rudolf (Hg.); Steuerung durch Indikatoren - methodologische und theoretische Reflektionen zur deutschen und internationalen Bildungsberichterstattung; Verlag Barbara Budrich, Opladen, S.17-30

ThILLM - Thüringer Institut für Lehrerfortbildung, Lehrplanentwicklung und Medien (ThILLM); Gehirngerechtes Klassenzimmer – „Handreichungen für die Unterrichtspraxis“; ThILLM-Heft 126

Weinert, Franz E. (2001a); Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit; in: Weinert, Franz E. (Hg.); Leistungsmessungen in Schulen; Beltz Verlag, Weinheim u.a., S. 17-31

Weinert, Franz E. (Hg.) (2001b); Leistungsmessungen in Schulen; Beltz Verlag, Weinheim u.a.

Weinert, Franz E. (2001c); Perspektiven der Schulleistungsmessung – mehrperspektivisch betrachtet; in: Weinert, Franz E. (Hg.); Leistungsmessungen in Schulen; Beltz Verlag, Weinheim u.a., S. 353-365

Wolter, Andrä (2009); Hochschulindikatoren in der nationalen Bildungsberichterstattung: Ihre Stärken und Schwächen; in: Tippelt, Rudolf (Hg.); Steuerung durch Indikatoren - methodologische und theoretische Reflektionen zur deutschen und internationalen Bildungsberichterstattung; Verlag Barbara Budrich, Opladen, S. 73-91

Wolter, Stefan C. (2008); Bildungsberichterstattung auf der Basis von Indikatoren; in: Landesinstitut für Schule und Medien Berlin-Brandenburg (Lisum) / Kunst und Kultur (bm:ukk, Österreich) / Bundesministerium für Unterricht (Schweiz) / Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK) (Hg.); Bildungsmonitoring, Vergleichsstudien und Innovationen: Von evidenzbasierter Steuerung zur Praxis; Berliner Wissenschafts-Verlag, Berlin, S. 53-70